

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:34:20 ; Search time 51.7 Seconds  
(without alignments)  
58.008 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155

Sequence: 1 EKCVMYTDGKWNDRNCLOSRLATCEF 27

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155	100.0	27	20	AAV25519 Human collectin pr
2	155	100.0	27	21	AAV77986 Modified consensus
3	155	100.0	27	22	AAV63349 Consensus sequence
4	137	88.4	117	7	AAV60439 3' Terminal portio
5	137	88.4	248	7	AAV60665 Sequence of human
6	137	88.4	248	7	AAV60666 Genomic sequence o
7	137	88.4	248	7	AAV60441 Plasmid pASPC-SV(1
8	137	88.4	248	7	AAV60442 35kd pulmonary sur
9	137	88.4	248	8	AAV70662 35kd pulmonary sur
10	137	88.4	248	8	AAV70663
11	137	88.4	248	9	AAV80694

12	137	88.4	248	9	AAV82980	Sequence deduced f
13	137	88.4	248	11	AAV05092	Product of vector
14	137	88.4	248	11	AAV05091	Vector PSP 35K-1A-
15	137	88.4	248	11	AAV04215	Human 32K ASP enco
16	137	88.4	248	11	AAV04216	Human 32K ASP enco
17	137	88.4	248	11	AAV06331	Human alveolar sur
18	137	88.4	248	21	AAV77989	Human SP-A amino a
19	137	88.4	271	7	AAV60661	Genomic sequence o
20	137	88.4	271	11	AAV04212	Human 32K alveolar
21	137	88.4	271	11	AAV04217	Human 32K ASP enco
22	135	87.1	259	21	AAV56135	Lung cancer associ
23	133	85.8	248	7	AAV60437	Dog 32 kd alveolar
24	133	85.8	256	11	AAV04210	Canine 32K alveola
25	108	69.7	375	21	AAV77990	Human SP-D amino a
26	95	61.3	351	18	AAV18780	Bovine conglutinin
27	95	61.3	351	18	AAV13672	Bovine conglutinin
28	95	61.3	371	16	AAV75642	Bovine conglutinin
29	84	54.2	249	21	AAV15401	Pig serum lectin.
30	84	54.2	335	12	AAV12222	Conglutinin. Bos
31	81	52.3	27	21	AAV77987	Consensus sequence
32	81	52.3	27	22	AAV63348	Consensus sequence
33	81	52.3	228	20	AAV03769	Mannan-binding pro
34	81	52.3	248	14	AAV45005	Human Mannose-Bind
35	81	52.3	248	20	AAV29485	Human mannan bindi
36	81	52.3	248	22	AAV36578	Human mannan bindi
37	81	52.3	277	20	AAV41698	Human collectin pr
38	81	52.3	277	20	AAV25518	Human collectin pr
39	81	52.3	277	21	AAV44254	Human PRO702 (UNQ3
40	81	52.3	277	22	AAV28073	Human PRO polypept
41	79.5	51.3	484	20	AAV99892	Porcine E-selectin
42	79.5	51.3	484	21	AAV81934	Porcine E-selectin
43	79.5	51.3	830	16	AAV65216	P-selectin. Homo
44	79	51.0	229	18	AAV15251	Asialoglycoprotein
45	79	51.0	270	18	AAV15252	Asialoglycoprotein

# ALIGNMENTS

## RESULT 1

AAV25519  
ID AAV25519 standard; Protein; 27 AA.

XX AC

AAV25519;

DT 30-SEP-1999 (first entry)

XX Human collectin protein consensus fragment.

DE Collectin: human; antibacterial; antiviral; treatment; infection.

XX Synthetic.

XX Homo sapiens.

PN W0937767-A1.

XX 29-JUL-1999.

XX 24-JUL-1998; 98WO-JP03328.

PF 23-JAN-1998; 98JP-0011281.

XX (FUSO ) FUSO PHARM IND LTD.

PA Wakamiya N;

PI WPI; 1999-458691/38.

XX New collectin protein of human origin and DNA encoding it

PT Claim 3; Page 44; 58pp; Japanese.

PS This invention describes the isolation and characterisation of a novel

CC

```
CC human collectin protein and its encoding polynucleotide. The human
CC collectin exhibits antibacterial and antiviral activity and can be used
CC as an agent for the treatment of human bacterial and viral infections.
CC This sequence represents a human collectin protein consensus sequence.
XX
XX
SQ Sequence 27 AA;

Query Match 100.0%; Score 155; DB 20; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-15;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCIQLSRIAICEF 27
Db 1 ekcvemytdgkwnrndncqlsrilaicef 27

RESULT 2
AAY77986
ID AAY77986 standard; peptide; 27 AA.
XX
AC AAY77986;
XX
XX 20-JUN-2000 (first entry)
XX
XX Modified consensus sequence of collectin.
XX
XX Collectin; human; antibacterial; antiviral.
XX
XX Homo sapiens.
XX
XX WO200011161-A1.
XX
XX 02-MAR-2000.
XX
XX 24-AUG-1999; 99WO-JP04552.
XX
XX 24-AUG-1998; 98JP-0237611.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI; 2000-224696/19.
XX
XX New collectin of human origin having antibacterial and antiviral
XX activity, and gene encoding it useful for production of transgenic
XX animals and of antibodies for screening potential drug molecules -
XX
XX Example 1; Page 31; 106pp; Japanese.
XX
XX The invention relates to polynucleotides encoding a new collectin of
XX human origin. The collectin can be used as an antibacterial and antiviral
XX agent and for screening potential drug molecules. The new collectin can
XX be produced by standard recombinant methodology. The present sequence
XX represents a modified consensus sequence of collectins hybridisable with
XX the novel human collectin.
XX
XX Sequence 27 AA;

Query Match 100.0%; Score 155; DB 21; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-15;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCIQLSRIAICEF 27
Db 1 ekcvemytdgkwnrndncqlsrilaicef 27

RESULT 3
AAG63349
ID AAG63349 standard; protein; 27 AA.
```

```
XX
AC AAG63349;
XX
DT 15-OCT-2001 (first entry)
XX
DE Consensus sequence of a collectin.
XX
XX Human; scavenger receptor; SRC1-P1; macrophage; basal immunity;
XX arteriosclerosis; diabetes; bacterial infection; restenosis; angioplasty;
XX low density lipoprotein; collectin.
XX
XX Synthetic.
XX
XX WO2001159107-A1.
XX
XX 16-AUG-2001.
XX
XX 08-FEB-2001; 2001WO-JP00874.
XX
XX 14-FEB-2000; 2000JP-0035155.
XX
XX 10-OCT-2000; 2000JP-0309068.
XX
XX (FUSO ) FUSO PHARM IND LTD.
XX
XX Wakamiya N;
XX
XX WPI; 2001-497076/54.
XX
XX New Scavenger receptor proteins SRC1-P1 with collectin-like structure,
XX useful for treatment and diagnosis of diseases associated with oxidized
XX low-density lipoprotein accumulation -
XX
XX Example 1; Page 47; 118pp; Japanese.
XX
XX The present sequence represents a collectin consensus sequence. The
XX specification describes human scavenger receptors designated. They
XX SRC1-P1. The SRC1-P1 polypeptide has a collectin-like structure. They
XX are useful in clarifying the functions of macrophages and basal
XX immunity. They are also useful in the treatment, prevention, diagnosis
XX and investigation of diseases such as arteriosclerosis, diabetic
XX complications, bacterial infection and restenosis following angioplasty,
XX which are associated with accumulation of oxidized low density
XX lipoprotein and the binding of advanced glycation end-products into
XX cells.
XX
XX Sequence 27 AA;

Query Match 100.0%; Score 155; DB 22; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-15;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCIQLSRIAICEF 27
Db 1 ekcvemytdgkwnrndncqlsrilaicef 27

RESULT 4
AAP60439
ID AAP60439 standard; protein; 117 AA.
XX
XX AAP60439;
XX
XX 01-JAN-1980 (first entry)
XX
XX 3' Terminal portion of human alveolar surfactant protein (ASP).
XX Alveolar surfactant protein; ASP; respiratory distress syndrome;
XX respiratory disease; pneumonia; bronchitis; ss.
XX
XX Homo sapiens.
XX
XX WO9603408-A.
```

```

XX PD 19-JUN-1986.
XX PF 10-DEC-1985; 85WO-US02445.
XX PR 11-DEC-1984; 84US-0680358.
XX PR 13-DEC-1985; 85US-0808843.
XX PR 29-JAN-1987; 87US-0008453.
XX PA (CALB-) CALIF BIOTECHNOL IN.
XX PI Schilling JW, White RT, Cordell B, Benson BJ;
DR WPI: 1986-169372/26.
DR N-PSDB; AAN60377.
XX Recombinant alveolar surfactant protein - for treatment of
PT respiratory distress syndrome and respiratory diseases
XX PS Disclosure; Fig. 5; 77pp; English.
XX CC The ASP can be used for the treatment of respiratory distress
CC syndrome in infants or adults and respiratory diseases such as
CC pneumonia and bronchitis. Preferably, the 32k component in
CC combination with the 10 k component of ASP is combined with
CC natural or synthetic lipids to construct a surfactant complex.
CC also AAN60375-6, AAN60378-87 and AAP60437-8, AAP60440-42 and
CC AAP60445-48.
XX SQ Sequence 117 AA;

Query Match 88.4%; Score 137; DB 7; Length 117;
Best Local Similarity 85.2%; Pred. No. 1e-11;
Matches 23; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
Db 91 eqcvemytdgqwndrnclysrliceef 117
:|||||:|||||:|||||:|||||

RESULT 5
AAP60665
ID AAP60665 standard; Protein; 248 AA.
XX AC AAP60665;
XX DT 31-JUL-1991 (first entry)
XX DE Sequence of human alveolar surfactant protein (hASP)
DE on pMT(E):HS and pASPC-SV(10).
XX KW Regulatable expression system.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein 21..248
XX PN W08604920-A.
XX PD 28-AUG-1986.
XX PF 11-FEB-1986; 86WO-US00296.
XX PR 25-NOV-1985; 85US-0801674.
XX PR 13-FEB-1985; 85US-0701296.
XX PA (BIOT-) BIOTECHN RES PARTNE.
XX PA (KUSH/) KUSHNER P J.
XX PI Kushner PJ, Cofer CL, Friedman J, Talmadge KD;

WPI: 1986-238888/36.
N-PSDB; AAN60572.
Regulatable expression systems - contg. human metallo:thionein-II
promoter
Example; Fig 6; 94pp; English.
A regulatable expression system for a coding sequence is claimed.
The system can process genomic as well as intronless DNA.
Sequence 248 AA;

Query Match 88.4%; Score 137; DB 7; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

XX WPI: 1986-238888/36.
XX DR N-PSDB; AAN60571.
XX Regulatable expression systems - contg. human metallo:thionein-II
PT promoter
XX PS Example; Fig 5; 94pp; English.
XX CC A regulatable expression system for a coding sequence is claimed.
CC The system can process genomic as well as intronless DNA.
XX SQ Sequence 248 AA;

Query Match 88.4%; Score 137; DB 7; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
Db 222 eqcvemytdgqwndrnclysrliceef 248
:|||||:|||||:|||||:|||||

RESULT 6
AAP60666
ID AAP60666 standard; Protein; 248 AA.
XX AC AAP60666;
XX DT 31-JUL-1991 (first entry)
XX DE Genomic sequence of human alveolar surfactant protein (hASP)
DE encoded by genomic DNA, used to obtain pASPCg-SV(10).
XX KW Regulatable expression system.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Peptide 1..20
FT Protein 21..248
XX PN W08604920-A.
XX PD 28-AUG-1986.
XX PF 11-FEB-1986; 86WO-US00296.
XX PR 25-NOV-1985; 85US-0801674.
XX PR 13-FEB-1985; 85US-0701296.
XX PA (BIOT-) BIOTECHN RES PARTNE.
XX PA (KUSH/) KUSHNER P J.
XX PI Kushner PJ, Cofer CL, Friedman J, Talmadge KD;

WPI: 1986-238888/36.
N-PSDB; AAN60572.
Regulatable expression systems - contg. human metallo:thionein-II
promoter
Example; Fig 6; 94pp; English.
A regulatable expression system for a coding sequence is claimed.
The system can process genomic as well as intronless DNA.
Sequence 248 AA;

Query Match 88.4%; Score 137; DB 7; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;

```

Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQSLRAICEF 27  
 |:|||||:|||||:|||||:|||||  
 Db 222 eqcvemytdgqwndrnclysrlicief 248

## RESULT 7

AAP60441  
 ID AAP60441 standard; protein; 248 AA.  
 XX  
 AC AAP60441;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE Plasmid pASpc-SV(10) encoding human alveolar surfactant  
 DE protein.  
 XX  
 KW Alveolar surfactant protein; ASP; canine; plasmid pASpc-SV(10);  
 KW respiratory distress syndrome; respiratory disease; pneumonia;  
 KW bronchitis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO8603408-A.  
 XX  
 PD 19-JUN-1986.  
 XX  
 PF 10-DEC-1985; 85WO-US02445.  
 XX  
 PR 11-DEC-1984; 84US-0680358.  
 PR 13-DEC-1985; 85US-0808843.  
 PR 29-JAN-1987; 87US-0008453.  
 XX  
 PA (CALB-) CALIF BIOTECHNOL IN.  
 XX  
 PI Schilling JW, White RT, Cordell B, Benson BJ;  
 XX  
 DR WPI: 1986-169372/26.  
 DR N-PSDB; AAN60379.  
 XX  
 PT Recombinant alveolar surfactant protein - for treatment of  
 PT respiratory distress syndrome and respiratory diseases  
 XX  
 PS Disclosure; Fig. 7; 77pp; English.  
 XX  
 CC The ASP can be used for the treatment of respiratory distress  
 CC syndrome in infants or adults and respiratory diseases such as  
 CC pneumonia and bronchitis. This plasmid is used for the expression  
 CC of human ASP in mammalian cell culture, e.g CHO cell culture. See  
 CC also AAN60375-78, AAN60380-87 and AAP60437-40, AAP60442-48.  
 XX  
 SQ Sequence 248 AA;

PI Schilling JW, White RT, Cordell B, Benson BJ;

DR WPI: 1986-169372/26.

DR N-PSDB; AAN60379.

PT Recombinant alveolar surfactant protein - for treatment of  
 PT respiratory distress syndrome and respiratory diseases

PS Disclosure; Fig. 7; 77pp; English.

CC The ASP can be used for the treatment of respiratory distress  
 CC syndrome in infants or adults and respiratory diseases such as  
 CC pneumonia and bronchitis. This plasmid is used for the expression  
 CC of human ASP in mammalian cell culture, e.g CHO cell culture. See  
 CC also AAN60375-78, AAN60380-87 and AAP60437-40, AAP60442-48.

XX Sequence 248 AA;

Query Match 88.4%; Score 137; DB 7; Length 248;  
 Best Local Similarity 85.2%; Pred. No. 2.2e-11;

Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQSLRAICEF 27  
 |:|||||:|||||:|||||:|||||  
 Db 222 eqcvemytdgqwndrnclysrlicief 248

## RESULT 8

AAP60442  
 ID AAP60442 standard; protein; 248 AA.  
 XX  
 AC AAP60442;  
 XX  
 DT 01-JAN-1980 (first entry)  
 XX  
 DE Plasmid pASpcq-SV(10) encoding human alveolar surfactant

DE protein.

XX Alveolar surfactant protein; ASP; canine; plasmid pASpcq-SV(10);  
 KW respiratory distress syndrome; respiratory disease; pneumonia;  
 KW bronchitis; ss.

XX OS Homo sapiens.

XX WO8603408-A.

XX PD 19-JUN-1986.

PF 10-DEC-1985; 85WO-US02445.

XX 11-DEC-1984; 84US-0680358.

PR 13-DEC-1985; 85US-0808843.

PR 29-JAN-1987; 87US-0008453.

XX (CALB-) CALIF BIOTECHNOL IN.

XX Schilling JW, White RT, Cordell B, Benson BJ;

XX WPI: 1986-169372/26.

DR N-PSDB; AAN60380.

PT Recombinant alveolar surfactant protein - for treatment of  
 PT respiratory distress syndrome and respiratory diseases

XX Disclosure; Fig. 8; 77pp; English.

XX The ASP can be used for the treatment of respiratory distress  
 CC syndrome in infants or adults and respiratory diseases such as  
 CC pneumonia and bronchitis. This plasmid is used for the expression  
 CC of human ASP in mammalian cell culture, e.g CHO cell culture. See  
 CC also AAN60375-79, AAN60381-87 and AAP60437-41, AAP60443-48.

XX Sequence 248 AA;

Query Match 88.4%; Score 137; DB 7; Length 248;

Best Local Similarity 85.2%; Pred. No. 2.2e-11;

Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQSLRAICEF 27  
 |:|||||:|||||:|||||:|||||  
 Db 222 eqcvemytdgqwndrnclysrlicief 248

## RESULT 9

AAP70662  
 ID AAP70662 standard; Protein; 248 AA.

XX AAP70662;

XX 29-APR-1991 (first entry)

DE 35kd pulmonary surfactant protein.

XX Hyaline membrane disease; respiratory distress syndrome; RDS.

XX OS Homo sapiens.

XX WO8702037-A.

XX 09-APR-1987.

XX 26-SEP-1986; 86WO-US02034.

XX 26-SEP-1985; 85US-0781130.

PR 15-AUG-1986; 86US-0897183.

XX (GENE-) GENETICS INSTITUTE INC.  
 PA (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.

```
XX
PI Tausch HW, Cobs KA, Steinbrink DR, Floros J, Phelps DS;
XX
XX WPI; 1987-108682/15.
DR N-PSDB; AAN71009.
XX
XX Pulmonary surfactant proteins - used for treating Hyaline Membrane
PT Disease or Respiratory Distress Syndrome.
XX
XX Claim 1; Page 33-34; 50pp; English.
XX
XX Gene product may be used in treatment of Hyaline Membrane Disease
CC and Respiratory Distress Syndrome (RDS) in both premature infants
CC and adults eg. cardio-pulmonary operations. The protein products
CC may also be used to raise diagnostic antibodies.
XX
XX Sequence 248 AA;
SQ

Query Match      88.4%; Score 137; DB 8; Length 248;
Best Local Similarity .85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdggwdrnclysrlicf 248
|:|||||:|||||:|||||:|||||

RESULT 10
AAP70663
ID AAP70663 standard; Protein; 248 AA.
XX
XX AAP70663;
AC
XX 29-APR-1991 (first entry)
DT
XX
XX 35kd pulmonary surfactant protein.
DE
XX Hyaline membrane disease; respiratory distress syndrome; RDS.
XX
XX Homo sapiens.
OS
XX WO8702037-A.
PN
XX 09-APR-1987.
PD
XX 26-SEP-1986; 86WO-US02034.
XX
XX 26-SEP-1985; 85US-0781130.
PR
XX 15-AUG-1986; 86US-0897183.
XX
XX (GENE-) GENETICS INSTITUTE INC.
PA
XX (BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.
XX
XX Tausch HW, Cobs KA, Steinbrink DR, Floros J, Phelps DS;
XX
XX WPI; 1987-108682/15.
DR N-PSDB; AAN71010.
XX
XX Pulmonary surfactant proteins - used for treating Hyaline Membrane
PT Disease or Respiratory Distress Syndrome.
XX
XX Claim 1; Page 34A-B; 50pp; English.
XX
XX Gene product may be used in treatment of Hyaline Membrane Disease
CC and Respiratory Distress Syndrome (RDS) in both premature infants
CC and adults eg. cardio-pulmonary operations. The protein products
CC may also be used to raise diagnostic antibodies.
XX
XX Sequence 248 AA;
SQ

Query Match      88.4%; Score 137; DB 8; Length 248;
```

```
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdggwdrnclysrlicf 248
|:|||||:|||||:|||||:|||||

RESULT 11
AAP80694
ID AAP80694 standard; protein; 248 AA.
XX
XX AAP80694;
AC
XX 04-DEC-1990 (first entry)
DT
XX
XX Sequence deduced from pHS10-5, encoding human 32K ASP.
DE
XX Alveolar surfactant protein; ASP; respiratory distress syndrome;
KW pneumonia; bronchitis; 32K.
XX Homo sapiens.
XX
XX WO8805820-A.
PN
XX 11-AUG-1988.
PD
XX 15-JAN-1987; 87WO-US00092.
XX
XX 29-JAN-1987; 87US-0008453.
PR
XX (CALB-) CALIF BIOTECHN INC.
PA
XX Schilling JW, White RT, Cordell B, Benson BJ;
PI
XX WPI; 1988-124493/18.
DR
XX Recombinant alveolar surfactant protein - used for treating
PT respiratory distress syndrome and related diseases e.g. pneumonia
PT and bronchitis.
XX
XX Disclosure; ; pp; English.
PS
XX
XX The sequence was deduced from clone pHS10-5, isolated from a
CC human lung cDNA library. The protein is part of the alveolar
CC surfactant protein, high mol. wt, hydrophilic 32k gp. The protein
CC differs at position 50 from a previously published (W08603408)
CC sequence, and differs at several places from two other sequences
CC determined by others. It is believed that the 32k human ASP
CC protein may be encoded by multiple genes. The recombinant protein
CC can be used for the treatment of respiratory disorders.
CC See also AAP82977-80, and AAP82982.
XX
XX Sequence 248 AA;
SQ

Query Match      88.4%; Score 137; DB 9; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.2e-11;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSRLAICEF 27
Db 222 eqcvemytdggwdrnclysrlicf 248
|:|||||:|||||:|||||:|||||

RESULT 12
AAP82980
ID AAP82980 standard; protein; 248 AA.
XX
XX AAP82980;
AC
XX 04-DEC-1990 (first entry)
DT
XX
```

DE Sequence deduced from pHS10-4, encoding human 32K ASP.  
XX  
KW Alveolar surfactant protein; ASP; respiratory distress syndrome;  
KW pneumonia; bronchitis; 32K.  
XX  
OS Homo sapiens.  
XX  
PN WO8005820-A.  
XX  
PD 11-AUG-1988.  
XX  
XX 15-JAN-1987; 87WO-US00092.  
XX  
XX 29-JAN-1987; 87US-0008453.  
XX  
XX (CALB-) CALIF BIOTECH INC.  
XX  
XX Schilling JW, White RT, Cordell B, Benson BJ;  
XX  
XX WPI; 1988-124493/18.  
XX  
XX Recombinant alveolar surfactant protein - used for treating  
PT respiratory distress syndrome and related diseases e.g. pneumonia  
PT and bronchitis.  
XX  
XX Disclosure: ; pp; English.  
XX  
XX The sequence was deduced from clone pHS10-4, isolated from a  
CC human lung cDNA library. The protein is part of the alveolar  
CC surfactant protein, high mol. wt, hydrophilic 32K gp. The protein  
CC differs at 7 positions from a previously published (W08603408)  
CC sequence, and also at several places from two other sequences  
CC determined by others. It is believed that the 32K human ASP  
CC protein may be encoded by multiple genes. The recombinant protein  
CC can be used for the treatment of respiratory disorders.  
CC See also AAF82977-79, AAF82982 and AAF80694.  
XX  
XX Sequence 248 AA;

Query Match 88.4%; Score 137; DB 9; Length 248;  
Best Local Similarity 85.2%; Pred. No. 2.2e-11;  
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27  
I:|||||:||||| ||| |||  
Db 222 eqcvemytdgqwndrnclyslrticef 248

RESULT 13  
AAR05092  
ID AAR05092 standard; protein; 248 AA.  
XX  
XX AAR05092;  
XX  
XX 11-JUL-1990 (first entry)  
XX  
DE Product of vector PSP 35K-68-A encoding pulmonary surfactant protein.  
XX  
XX Respiratory distress syndrome; RDS; hyaline membrane disease;  
KW pulmonary surfactant protein;  
XX  
XX Homo sapiens.  
XX  
XX US4882422-A.  
XX  
XX 21-NOV-1989.  
XX  
XX 24-SEP-1987; 87US-0100372.  
XX  
XX 26-SEP-1985; 85US-0791120.  
PR 15-AUG-1986; 86US-0897183.  
XX

PA (GENE-) GENETICS INSTITUTE.  
XX  
XX Taush HW, Jacobs KA, Steinbrink DR, Floros J, Phelps DS, Fritsch EF;  
XX  
XX WPI; 1990-036829/05.  
DR N-PSDB; AAQ02893.  
XX  
XX Purified human pulmonary surfactant protein -  
PT useful for treating respiratory distress syndrome.  
PT  
XX  
XX Disclosure; Table 2; 15pp; English.  
XX  
XX Protein is useful in treatment of respiratory distress syndrome (Hyaline  
CC membrane disease) enhancing pulmonary surfactant activity.  
CC Protein is encoded by the cDNA insert in vector PSP 35K-68-A  
CC (ATCC 40244).  
XX  
XX Sequence 248 AA;

Query Match 88.4%; Score 137; DB 11; Length 248;  
Best Local Similarity 85.2%; Pred. No. 2.2e-11;  
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27  
I:|||||:||||| ||| |||  
Db 222 eqcvemytdgqwndrnclyslrticef 248

RESULT 14  
AAR05091  
ID AAR05091 standard; protein; 248 AA.  
XX  
XX AAR05091;  
XX  
XX 11-JUL-1990 (first entry)  
XX  
XX Vector PSP 35K-1A-10 gene product encoding pulmonary surfactant protein.  
DE  
XX Respiratory distress syndrome; RDS; hyaline membrane  
KW disease; pulmonary surfactant protein;  
XX  
XX Homo sapiens.  
XX  
XX US4882422-A.  
PN  
XX 21-NOV-1989.  
PD  
XX 24-SEP-1987; 87US-0100372.  
PF  
XX 26-SEP-1985; 85US-0791120.  
PR 15-AUG-1986; 86US-0897183.  
XX  
XX (GENE-) GENETICS INSTITUTE.  
XX  
XX Taush HW, Jacobs KA, Steinbrink DR, Floros J, Phelps DS, Fritsch EF;  
PI  
XX WPI; 1990-036829/05.  
DR N-PSDB; AAQ02892.  
XX  
XX Purified human pulmonary surfactant protein -  
PT useful for treating respiratory distress syndrome.  
PT  
XX  
XX Disclosure; Table 1; 15pp; English.  
XX  
XX Protein is useful in treatment of respiratory distress syndrome (Hyaline  
CC membrane disease) enhancing pulmonary surfactant activity.  
CC Protein is encoded by the cDNA insert in vector PSP 35K-1A-10  
CC (ATCC 40243).  
XX  
XX Sequence 248 AA;

Query Match 88.4%; Score 137; DB 11; Length 248;  
Best Local Similarity 85.2%; Pred. No. 2.2e-11;  
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVENYTDGKWNDRNCLOSRLAICEF 27  
|:|||||:||||| ||| |||  
Db 222 egcvemytdgqwndrnclysrliceef 248

## RESULT 15

AA04215  
ID AAR04215 standard; protein; 248 AA.

XX AC AAR04215;

XX DT 12-SEP-1989 (first entry)

XX DE Human 32K ASP encoded by pASpc-SV(10).

XX KW Mammalian 32K alveolar surfactant protein; respiratory distress syndrome;  
XX lungs.

XX OS synthetic.

XX PN US4912038-A.

XX PD 27-MAR-1990.

XX PF 13-DEC-1985; 85US-0808843.

XX PR 11-DEC-1984; 84US-0680358.

XX PA (CALB-) CALIFORNIA BIOTECHNOLOGY INC.

XX PI Schilling JW, White RT, Cordell B, Benson RJ;

XX DR WPI; 1990-139399/18.

XX DR N-PSDB; AAQ04080.

XX PT Recombinant vector contg. DNA encoding 32K alveolar surfactant  
XX protein used for treatment of eg respiratory distress syndrome

XX PS Claim 1; Fig 7; 70pp; English.

XX CC This sequence is the human 32K alveolar surfactant protein which is  
XX encoded by the expression vector pASpc-SV(10) which is used to transform  
XX mammalian host cells (eg CHO cells) resulting in efficient expression  
XX of the protein.

XX CC See also AAR04210-12, AAR04214 and R04216-17.

XX SQ Sequence 248 AA;

Query Match 88.4%; Score 137; DB 11; Length 248;  
Best Local Similarity 85.2%; Pred. No. 2.2e-11;  
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVENYTDGKWNDRNCLOSRLAICEF 27  
|:|||||:||||| ||| |||  
Db 222 egcvemytdgqwndrnclysrliceef 248

Search completed: July 3, 2002, 12:34:20  
Job time: 374 sec

\_\_\_\_\_

\_\_\_\_\_



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:34:47 ; Search time 20.65 Seconds  
(without alignments)  
31.937 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155

Sequence: 1 EKCVMYTDGKWNDRNCLOSLAICEF 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*

1: /cgn2.6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2.6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2.6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2.6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2.6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2.6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	84.5	111	6	Patent No. 5514582
2	95	61.3	161	3	Sequence 6, Appl
3	95	61.3	351	3	US-09-011-735-6
4	81.5	52.6	830	5	PCT-US91-05059-2
5	81.5	52.6	830	6	PCT-US91-05059-2
6	81	52.3	248	4	Sequence 2, Appl
7	80.5	51.9	119	1	US-09-198-603C-2
8	79.5	51.3	117	1	US-08-340-539A-16
9	79.5	51.3	119	1	US-08-274-661B-39
10	79.5	51.3	484	2	US-08-340-539A-17
11	79.5	51.3	484	3	US-08-252-493C-9
12	79.5	51.3	830	1	US-09-276-197-9
13	79	51.0	287	4	US-08-110-158-4
14	74	47.7	114	6	US-09-111-470-6
15	72.5	46.8	119	1	Patent No. 5514582
16	72.5	46.8	120	1	Sequence 15, Appl
17	72.5	46.8	120	1	US-08-340-539A-15
18	72.5	46.8	120	1	US-08-274-661B-36
19	72.5	46.8	610	3	US-08-274-661B-37
20	72.5	46.8	610	3	US-08-365-470-3
21	72.5	46.8	610	3	US-09-209-668-19
22	72.5	46.8	610	4	Sequence 89, Appl
23	72.5	46.8	610	6	US-09-009-490A-89
24	72.5	46.8	610	6	US-09-009-490A-89
25	72.5	46.8	610	6	US-09-009-490A-89
26	72.5	46.8	610	6	US-09-009-490A-89
27	72.5	46.8	610	6	US-09-009-490A-89
28	72.5	46.8	610	6	US-09-009-490A-89
29	72.5	46.8	610	6	US-09-009-490A-89
30	72.5	46.8	610	6	US-09-009-490A-89
31	72.5	46.8	610	6	US-09-009-490A-89
32	72.5	46.8	610	6	US-09-009-490A-89
33	72.5	46.8	610	6	US-09-009-490A-89
34	72.5	46.8	610	6	US-09-009-490A-89
35	72.5	46.8	610	6	US-09-009-490A-89
36	72.5	46.8	610	6	US-09-009-490A-89
37	72.5	46.8	610	6	US-09-009-490A-89
38	72.5	46.8	610	6	US-09-009-490A-89
39	72.5	46.8	610	6	US-09-009-490A-89
40	72.5	46.8	610	6	US-09-009-490A-89
41	72.5	46.8	610	6	US-09-009-490A-89
42	72.5	46.8	610	6	US-09-009-490A-89
43	72.5	46.8	610	6	US-09-009-490A-89
44	72.5	46.8	610	6	US-09-009-490A-89
45	72.5	46.8	610	6	US-09-009-490A-89

28 68.5 44.2 119 1 US-08-340-539A-13  
29 68.5 44.2 372 2 US-08-513-278-4  
30 68.5 44.2 372 6 5514582-4  
31 67.5 43.5 119 1 US-08-340-539A-19  
32 66.5 42.9 117 1 US-08-274-661B-38  
33 66.5 42.9 372 2 US-08-513-278-2  
34 66.5 42.9 372 6 5514582-2  
35 66.5 42.9 385 1 US-08-340-539A-2  
36 66.5 42.9 385 2 US-08-461-592B-2  
37 66.5 42.9 1455 3 US-08-840-062-5  
38 66 42.6 107 6 5514582-17  
39 66 42.6 108 6 5514582-16  
40 62.5 40.3 1479 3 US-08-840-062-2  
41 62.5 40.3 1479 3 US-08-840-062-4  
42 62 40.0 114 6 5514582-13  
43 62 40.0 114 6 5514582-14  
44 62 40.0 273 4 US-09-111-470-10  
45 62 40.0 292 2 US-08-688-342-4

#### ALIGNMENTS

RESULT 1  
5514582-20  
; Patent No. 5514582  
; APPLICANT: CAPON, DANIEL J.; LASKY, LAURENCE A.  
; TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID  
; IMMUNOGLOBULINS  
; NUMBER OF SEQUENCES: 43  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/185,670  
; FILING DATE: 21-JAN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 986,931  
; FILING DATE: 08-DEC-1992  
; APPLICATION NUMBER: 808,122  
; FILING DATE: 16-DEC-1991  
; APPLICATION NUMBER: 440,625  
; FILING DATE: 22-NOV-1989  
; APPLICATION NUMBER: 315,015  
; FILING DATE: 23-FEB-1989  
; SEQ ID NO: 20:  
; LENGTH: 111  
5514582-20

Query Match 84.5%; Score 131; DB 6; Length 111;  
Best Local Similarity 88.0%; Pred. No. 7.1e-12;  
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAIC 25  
|||||  
DB 87 EKCVMYTDGKWNDRNCLOSLAIC 111

RESULT 2  
US-09-011-735-6  
; Sequence 6, Application US/09011735B  
; Patent No. 6110708  
; GENERAL INFORMATION:  
; APPLICANT: Wakamiya, No. 6110708utaka  
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof  
; FILE REFERENCE: 19036/34548  
; CURRENT APPLICATION NUMBER: US/09/011,735B  
; CURRENT FILING DATE: 1998-05-22  
; EARLIER APPLICATION NUMBER: JP 7-209698  
; EARLIER FILING DATE: 1995-08-17  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 161  
; TYPE: PRT

```
; ORGANISM: Bovine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (2)
; OTHER INFORMATION: Xaa is a protein-constituting amino acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3)
; OTHER INFORMATION: Xaa is a protein-constituting amino acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (5)
; OTHER INFORMATION: Xaa is a protein-constituting amino acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (6)
; OTHER INFORMATION: Xaa is a protein-constituting amino acid
US-09-011-735-6

Query Match 61.3%; Score 95; DB 3; Length 161;
Best Local Similarity 59.3%; Pred. No. 1.5e-06;
Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDNCLOSLAICEF 27
Db 135 ENCVEIFPDGKNDVPCSKQLLVICF 161

RESULT 3
US-09-011-735-1
; Sequence 1, Application US/09011735B
; Patent No. 6110708
; GENERAL INFORMATION:
; APPLICANT: Wakamaya, No. 6110708utaka
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof
; FILE REFERENCE: 19036/34548
; CURRENT APPLICATION NUMBER: US/09/011,735B
; EARLIER FILING DATE: 1998-05-22
; EARLIER FILING DATE: 1995-08-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Bovine
US-09-011-735-1

Query Match 61.3%; Score 95; DB 3; Length 351;
Best Local Similarity 59.3%; Pred. No. 3.5e-06;
Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDNCLOSLAICEF 27
Db 325 ENCVEIFPDGKNDVPCSKQLLVICF 351

RESULT 4
PCT-US91-05059-2
; Sequence 2, Application PC/TUS9105059
; GENERAL INFORMATION:
; APPLICANT: Regents of the Board of the, University of
; APPLICANT: Oklahoma
; TITLE OF INVENTION: Functionally Active Selectin-Derived
; TITLE OF INVENTION: Peptides
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street, Suite 3100
; CITY: Atlanta
; STATE: Georgia
```

```
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/05059
; FILING DATE: 19910717
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/320408
; FILING DATE: 08-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/554199
; FILING DATE: 17-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Padst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMRF110CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 830 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: YES
; ANTI-SENSE: YES
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Blood
; CELL TYPE: Endothelial
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 4..25
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 60..158
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 131..150
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 163..174
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 168..183
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 185..194
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 200..244
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 213..226
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 230..257
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 262..306
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 275..288
; NAME/KEY: Disulfide-bond
```



Db 129 EDCVEIYIKSPAGKWNDEMCKLKKMALC 158

## RESULT 6

US-09-198-603C-2  
; Sequence 2, Application US/09198603C

; Patent No. 6337193

; GENERAL INFORMATION:

; APPLICANT: TULLY, Raymond E.

; APPLICANT: CALTAGIRONE, G. Thomas

; APPLICANT: MOYER, Shawn S.

; APPLICANT: KONNING, Michael T.

; TITLE OF INVENTION: EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC

; FILE REFERENCE: YEAST

; CURRENT APPLICATION NUMBER: A7290

; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 248

; TYPE: PRT

; ORGANISM: Human

US-09-198-603C-2

Query Match 52.3%; Score 81; DB 4; Length 248;

Best Local Similarity 51.9%; Pred. No. 0.00025;

Matches 14; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCIQLSLAICEF 27

|||||:|||||:|||||:|||||

Db 220 EDCVLLKNGQWMDVPCSTSLAYCEF 246

## RESULT 7

US-08-340-539A-16

; Sequence 16, Application US/08340539A

; Patent No. 5808025

; GENERAL INFORMATION:

; APPLICANT: Tedder, Thomas F.

; APPLICANT: Kansas, Geoffrey S.

; TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FISH & NEAVE

; STREET: 1251 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10020

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/340,539A

; FILING DATE: 16-NOV-1994

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/008,459

; FILING DATE: 25-JAN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Gunnison, Jane

; REGISTRATION NUMBER: 38,479

; REFERENCE/DOCKET NUMBER: CG-104 CON

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-596-9000

; TELEFAX: 212-596-9090

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 119 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-340-539A-16

Query Match

51.9%; Score 80.5; DB 1; Length 119;

Best Local Similarity 46.7%; Pred. No. 0.00013;

Matches 14; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

QY 1 EKCVMY-----TDGKWNDRNCIQLSLAIC 25

|||||:|||||:|||||:|||||

Db 88 EDCVEIYIKREKDSGKWNDEKCTKQLALC 117

## RESULT 8

US-08-274-661B-39

; Sequence 39, Application US/08274661B

; Patent No. 5593882

; GENERAL INFORMATION:

; APPLICANT: Erbe, David V.

; APPLICANT: Lasky, Laurence A.

; APPLICANT: Presta, Leonard G.

; TITLE OF INVENTION: Selectin Variants

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Winpatin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/274,661B

; FILING DATE: 13-Jul-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/956701

; FILING DATE: 10/01/1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Dreger, Ginger R.

; REGISTRATION NUMBER: 33,055

; REFERENCE/DOCKET NUMBER: 761P1C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-3216

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

; INFORMATION FOR SEQ ID NO: 39:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 117 amino acids

; TYPE: Amino Acid

; TOPOLOGY: Linear

US-08-274-661B-39

Query Match 51.3%; Score 79.5; DB 1; Length 117;

Best Local Similarity 46.7%; Pred. No. 0.00018;

Matches 14; Conservative 5; Mismatches 6; Indels 5; Gaps 1;

QY 1 EKCVMY-----TDGKWNDRNCIQLSLAIC 25

|||||:|||||:|||||:|||||

Db 88 EDCVEIYIKSPAGKWNDEKCKLKKHALC 117

## RESULT 9

US-08-340-539A-17

```
; Sequence 17, Application US/08340539A
; Patent No. 5808025
; GENERAL INFORMATION:
; APPLICANT: Tedder, Thomas F.
; APPLICANT: Kansas, Geoffrey S.
; TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMUL/PANEOUS
; TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,539A
; FILING DATE: 16-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/008,459
; FILING DATE: 25-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gunnison, Jane
; REGISTRATION NUMBER: 38,479
; REFERENCE/DOCKET NUMBER: CC-104 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 119 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-340-539A-17

Query Match 51.3%; Score 79.5; DB 1; Length 119;
Best Local Similarity 46.7%; Pred. No. 0.00019;
Matches 14; Conservative 5; Mismatches 6; Indels 5; Gaps 1;

Qy 1 EKCVMY-----TDGKWNDRNCQSRLAIC 25
| ||||| | : |||
Db 88 EDCVEIYKSPAGKWNDEHCLKKKHALC 117

RESULT 10
US-08-252-493C-9
; Sequence 9, Application US/08252493C
; Patent No. 5891645
; GENERAL INFORMATION:
; APPLICANT: Rollins, Scott
; APPLICANT: Rother, Russell P.
; APPLICANT: Evans, Mark J.
; APPLICANT: Matlis, Louis A.
; TITLE OF INVENTION: PORCINE E-SELECTIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seth A. Fidel
; STREET: 25 Science Park, Box 15
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06511
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 750 Kb storage
; COMPUTER: PC compatible
; OPERATING SYSTEM: DOS 6.2
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,197
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/252,493
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fidel, Seth A.
; REGISTRATION NUMBER: 38,449
; REFERENCE/DOCKET NUMBER: ALX-138
; TELECOMMUNICATION INFORMATION:
```

```
; COMPUTER: PC compatible
; OPERATING SYSTEM: DOS 6.2
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,493C
; FILING DATE: June 1, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fidel, Seth A.
; REGISTRATION NUMBER: 38,449
; REFERENCE/DOCKET NUMBER: ALX-138
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 776-1790
; TELEFAX: (203) 772-3655
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 484 amino acids
; TYPE: amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: predicted amino acid sequence of
; DESCRIPTION: Porcine E-selectin
; US-08-252-493C-9

Query Match 51.3%; Score 79.5; DB 2; Length 484;
Best Local Similarity 46.7%; Pred. No. 0.00083;
Matches 14; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

Qy 1 EKCVMY-----TDGKWNDRNCQSRLAIC 25
| ||||| | : |||
Db 110 EDCVEIYKRDGSKWNDERCKSKKALC 139

RESULT 11
US-09-276-197-9
; Sequence 9, Application US/09276197
; Patent No. 6040428
; GENERAL INFORMATION:
; APPLICANT: Rollins, Scott
; APPLICANT: Rother, Russell P.
; APPLICANT: Evans, Mark J.
; APPLICANT: Matlis, Louis A.
; TITLE OF INVENTION: PORCINE E-SELECTIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seth A. Fidel
; STREET: 25 Science Park, Box 15
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06511
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 750 Kb storage
; COMPUTER: PC compatible
; OPERATING SYSTEM: DOS 6.2
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/276,197
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/252,493
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fidel, Seth A.
; REGISTRATION NUMBER: 38,449
; REFERENCE/DOCKET NUMBER: ALX-138
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: (203) 776-1790  
TELEFAX: (203) 772-3655  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 484 amino acids  
TYPE: amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA to mRNA  
DESCRIPTION: predicted amino acid sequence of  
DESCRIPTION: Porcine E-selectin  
US-09-276-197-9

Query Match 51.3%; Score 79.5; DB 3; Length 484;  
Best Local Similarity 46.7%; Pred. No. 0.00083;  
Matches 14; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

QY 1 EKCVMY-----TDGKWNDRNCQSRLAIC 25  
| |||:| |||| | : |||  
DB 110 EDCVEIYIKRDKSGKWNDRGSKKKLALC 139

RESULT 12  
US-08-110-158-4  
; Sequence 4, Application US/08/110158  
; Patent No. 5605821  
; GENERAL INFORMATION:  
; APPLICANT: McEever, Rodger P.  
; APPLICANT: Pan, Junliang  
; TITLE OF INVENTION: Expression Control Sequences of the  
; TITLE OF INVENTION: P-Selectin Gene  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 1100 Peachtree Street, Suite 2800  
; CITY: Atlanta  
; STATE: GA  
; COUNTRY: USA  
; ZIP: 30309-4530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/110,158  
; FILING DATE: 19930820  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/320,408  
; FILING DATE: 08-MAR-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404)-815-6508  
; TELEFAX: (404)-815-6555  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 830 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-110-158-4

Query Match 51.3%; Score 79.5; DB 1; Length 830;  
Best Local Similarity 46.7%; Pred. No. 0.0015;  
Matches 14; Conservative 5; Mismatches 6; Indels 5; Gaps 1;

QY 1 EKCVMY-----TDGKWNDRNCQSRLAIC 25

DB 129 EDCVEIYIKSPAGKWNDRHCLKKHALC 158  
| |||:| |||| | : |||

RESULT 13  
US-09-111-470-6  
; Sequence 6, Application US/09/111470  
; Patent No. 6277959  
; GENERAL INFORMATION:  
; APPLICANT: Valladeau, Jenny  
; APPLICANT: Ravel, Odile  
; APPLICANT: Bates, Elizabeth E.M.  
; APPLICANT: Ford, John  
; APPLICANT: Saeland, Sem  
; APPLICANT: Lebecque, Serge J.E.  
; TITLE OF INVENTION: Mammalian Membrane Protein Genes;  
; TITLE OF INVENTION: Related Reagents  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/111,470  
; FILING DATE: 08-JUL-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/053,080  
; FILING DATE: 09-JUL-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: SF0695  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650)852-9196  
; TELEFAX: (650)496-1200  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 287 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-111-470-6

Query Match 51.0%; Score 79; DB 4; Length 287;  
Best Local Similarity 53.8%; Pred. No. 0.00056;  
Matches 14; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQSRLAICE 26  
| ||||:| |||| | : |||

DB 252 EDCVEVQPDGRWNDRFCLQVIRWVCE 277

RESULT 14  
5514582-8  
; Patent No. 5514582  
; APPLICANT: CAPON, DANIEL J.; LASKY, LAURENCE A.  
; TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID  
; IMMUNOGLOBULINS  
; NUMBER OF SEQUENCES: 43  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/185,670  
; FILING DATE: 21-JAN-1994



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:40:21 ; Search time 42.99 Seconds  
(without alignments)  
108.650 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155  
Sequence: 1 EKCVMYTDGKNDRNCQLQSLAICEF 27

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL19.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rhodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	140	90.3	248	6 Q9N0G1	Q9n0g1 equus cabal
2	140	90.3	248	6 Q95L88	Q95l88 equus cabal
3	140	90.3	248	11 Q9CQI1	Q9cqil mus musculus
4	139	89.7	116	6 Q9NIX3	Q9nix3 sus scrofa
5	138	89.0	248	6 Q9T0G6	Q9tto6 ovis aries
6	110	71.0	222	13 Q90XB2	Q90xb2 gallus gall
7	99	63.9	378	6 Q9NIX4	Q9nix4 sus scrofa
8	95	61.3	335	6 Q97748	Q97748 bos taurus
9	85	54.8	236	6 Q28518	Q28518 macaca mula
10	85	54.8	240	6 Q9XSW3	Q9xsw3 sus scrofa
11	83.5	53.9	610	6 Q95LG1	Q95lg1 equus cabal
12	82.5	53.2	484	6 Q95LG2	Q95lg2 ovis aries
13	81	52.3	248	4 Q96TF9	Q96tf9 homo sapien
14	81	52.3	248	4 Q96TF8	Q96tf8 homo sapien
15	81	52.3	248	4 Q96TF7	Q96tf7 homo sapien
16	81	52.3	248	4 Q96KE4	Q96ke4 homo sapien

17	81	52.3	277	4 Q9Y6Z7	Q9y6z7 homo sapien
18	80	51.6	245	6 Q28517	Q28517 macaca mula
19	79.5	51.3	482	6 Q28982	Q28982 sus scrofa
20	79.5	51.3	485	6 Q95LG3	Q95lg3 odocoileus
21	79.5	51.3	616	4 Q95509	Q95509 homo sapien
22	79.5	51.3	740	4 Q95507	Q95507 homo sapien
23	79.5	51.3	740	4 Q95508	Q95508 homo sapien
24	79	51.0	287	4 Q03969	Q03969 homo sapien
25	79	51.0	292	4 Q00448	Q00448 homo sapien
26	77.5	50.0	609	6 Q9GLF0	Q9glf0 canis famil
27	77	49.7	224	11 Q9Z294	Q9z294 rattus sp.
28	74.5	48.1	646	6 Q29097	Q29097 sus scrofa
29	72.5	46.8	376	6 Q28629	Q28629 oryctolagus
30	71	45.8	271	4 Q9BWP8	Q9bwp8 homo sapien
31	69	44.5	272	11 Q9DC75	Q9dc75 mus musculus
32	68.5	44.2	372	11 Q63762	Q63762 rattus norv
33	67.5	43.5	649	6 Q28657	Q28657 oryctolagus
34	67.5	43.5	754	6 Q28290	Q28290 canis famil
35	66.5	42.9	385	4 Q9UJ43	Q9uj43 homo sapien
36	66.5	42.9	1456	11 Q61830	Q61830 mus musculus
37	65	41.9	196	11 Q9EPW4	Q9epw4 mus musculus
38	65	41.9	197	4 Q75596	Q75596 homo sapien
39	65	41.9	284	11 Q91V84	Q91v84 mus musculus
40	62.5	40.3	1479	4 Q9UBG0	Q9ubg0 homo sapien
41	62.5	40.3	1479	4 Q9Y5P9	Q9y5p9 homo sapien
42	62.5	40.3	1479	11 Q64449	Q64449 mus musculus
43	62	40.0	292	4 Q14538	Q14538 homo sapien
44	61.5	39.7	238	13 Q57451	Q57451 gallus gall
45	61.5	39.7	254	13 Q98TA4	Q98ta4 gallus gall

#### ALIGNMENTS

RESULT 1  
Q9N0G1  
ID Q9N0G1 PRELIMINARY; PRT: 248 AA.  
AC Q9N0G1;  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE PULMONARY SURFACTANT PROTEIN A.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=THOROUGHRED; TISSUE=LUNG;  
RA Hobo S.;  
RT "Molecular cloning of equine pulmonary surfactant proteins.";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB015963; BAA97976.1; -;  
DR HSSP: P35247; 1808.  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR001304; lectin\_c.  
DR Pfam: PF00059; lectin\_c; 1.  
DR SMART: SM00034; CLECT; 1.  
DR PROSITE: PS00615; C\_TYPE\_LLECTIN\_1; 1.  
DR PROSITE: PS00041; C\_TYPE\_LLECTIN\_2; 1.  
SQ SEQUENCE 248 AA; 26000 MW; BBE12EFB05C2B8D1 CRC64;

Query Match 90.3%; Score 140; DB 6; Length 248;  
Best Local Similarity 88.9%; Pred. No. 1e-13;  
Matches 24; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRNCQLQSLAICEF 27  
|||||  
DB 222 EKCVMYTDGKNDRNCQLQSLAICEF 248

RESULT 2



```

Q95L88
ID Q95L88 PRELIMINARY; PRT; 248 AA.
AC Q95L88;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN A.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RL Weber B.I.L., Hospes R., Gortner L.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF400580; AAL07690.1; -.
SQ SEQUENCE 248 AA; 26047 MW; B71133E005C9A5C1 CRC64;

Query Match 90.3%; Score 140; DB 6; Length 248;
Best Local Similarity 88.9%; Pred. No. 1e-13;
Matches 24; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27
Db 222 EKCVEMTDGKWNDRNCLOSLRAICEF 248

RESULT 3
Q9CQI1 PRELIMINARY; PRT; 248 AA.
ID Q9CQI1;
AC Q9CQI1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SURFACTANT ASSOCIATED PROTEIN A.
GN SFTPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO, AND LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Willming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AB011333; BAB27551.1; -.
DR EMBL: AK004620; BAB23416.1; -.
DR EMBL: AK004789; BAB23565.1; -.
DR HSSP: P35247; 1B08.
DR MGD: MGI:109518; Sftpa.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; Lectin_c.
DR Pfam: PF00059; lectin_c; 1.

```

```

DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 248 AA; 26183 MW; 8A5670CFAD3EB9B6 CRC64;

Query Match 90.3%; Score 140; DB 11; Length 248;
Best Local Similarity 88.9%; Pred. No. 1e-13;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27
Db 222 EKCVEMTDGKWNDRNCLOSLRAICEF 248

RESULT 4
Q9NIX3 PRELIMINARY; PRT; 116 AA.
ID Q9NIX3;
AC Q9NIX3;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE LUNG SURFACTANT PROTEIN A (FRAGMENT).
GN SFTPA.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20109098; PubMed=10640760;
RA van Eljik M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
RA Lawson P.R.;
RT "Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
RT localisation and tissue distribution.";
RL J. Immunol. 164:1442-1450(2000).
DR EMBL: AF133668; AAF28384.1; -.
DR HSSP: P35247; 1B08.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
FT NON_TER 1
FT SEQUENCE 116 AA; 12946 MW; 6601D39320760C74 CRC64;

Query Match 89.7%; Score 139; DB 6; Length 116;
Best Local Similarity 88.9%; Pred. No. 6.4e-14;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27
Db 90 EKCVEMTDGKWNDRNCLOSLRAICEF 116

RESULT 5
Q9TT06 PRELIMINARY; PRT; 248 AA.
ID Q9TT06;
AC Q9TT06;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PULMONARY SURFACTANT PROTEIN A (PULMONARY SURFACTANT-ASSOCIATED
DE PROTEIN A).
GN SPAS OR SP-A.
OC Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.

```

```
RX MEDLINE=20215262; PubMed=10749753;
RA Braems G.A., Yao L.-J., Inchley K., Brickenden A., Han V.K.M.,
RA Grolla A., Challis J.R.G., Rossmayer P.:
RT "Ovine surfactant protein cDNAs: use in studies on fetal lung growth
RT and maturation after prolonged hypoxemia.";
RL Am. J. Physiol. 278:L754-L764(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=20215263; PubMed=10749754;
RA Pietschmann S.M., Pison U.;
RT "cDNA cloning of ovine pulmonary SP-A, SP-B, and SP-C: isolation of
RT two different sequences for SP-B.";
RL Am. J. Physiol. 278:L765-L778(2000).
DR EMBL: AF211856; AAF18995.1; -
DR EMBL: AF076633; AAF31148.1; -
DR HSP: P19999; IYTT.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 1.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLEC; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
SQ SEQUENCE 248 AA; 26394 MW; D65E7293BBFF1FD9 CRC64;

Query Match 89.0%; Score 138; DB 6; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.1e-13;
Matches 23; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLQSLAICEF 27
   |||||:||||:||||:||||:||||:
Db 222 EKCVMYTDGWNKNCQLQSLAICEF 248

RESULT 6
ID Q90XB2 PRELIMINARY; PRT; 222 AA.
AC Q90XB2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SURFACTANT PROTEIN A PRECURSOR.
GN SP-A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Vitved L., Hansen S., Teisner B., Koch C., Juul-Madsen H.,
RA Holmskov U., Salomonsen J., Skjoldt K.;
RT "The chicken homolog of surfactant protein A has only three collagen-
RT like Gly-Xaa-Yaa repeats and contain a new putative coil structure
RT between the collagen region and the alpha-helical coil-coil region.";
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF411083; AAK97540.1; -
DR Signal.
FT SIGNAL 1 18 POTENTIAL.
SQ SEQUENCE 222 AA; 24721 MW; 97371A3C53303872 CRC64;

Query Match 71.0%; Score 110; DB 13; Length 222;
Best Local Similarity 66.7%; Pred. No. 3.9e-09;
Matches 18; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLQSLAICEF 27
   |||||:||||:||||:||||:||||:
Db 196 EKCVMYTDGWNKNCQLQSLAICEF 222
```

```
RESULT 7
ID Q9N1X4 PRELIMINARY; PRT; 378 AA.
AC Q9N1X4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LUNG SURFACTANT PROTEIN D PRECURSOR.
GN SFTPD.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20109098; PubMed=10640760;
RA van Eljk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
RA Lawson P.R.;
RT "Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
RT localisation and tissue distribution.";
RL J. Immunol. 164:1442-1450(2000).
DR EMBL: AF132496; AAF22145.2; -
DR HSP: P35247; 1B08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 3.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLEC; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 378 LONG SURFACTANT PROTEIN D.
SQ SEQUENCE 378 AA; 37986 MW; 3504B8C1E56C341D CRC64;

Query Match 63.9%; Score 99; DB 6; Length 378;
Best Local Similarity 59.3%; Pred. No. 3.5e-07;
Matches 16; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCQLQSLAICEF 27
   |||||:||||:||||:||||:||||:
Db 352 ENCVEIFPGKWNKACGELRLVICEF 378

RESULT 8
ID O97748 PRELIMINARY; PRT; 335 AA.
AC O97748;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CONGLUTININ.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93277452; PubMed=7684896;
RA Lu J., Laursen S., Thiel S., Jensenius J., Reid K.;
RT "The cDNA cloning of conglutinin and identification of liver as a
RT primary site of synthesis of conglutinin in members of the Bovidae.";
```

	Query Match	Score 85:	DB 6:	Length 236;	
	Best Local Similarity	55.6%;	Pred. No. 3e-05;		
	Matches 15: Conservative	3:	Mismatches 9:	Indels 0:	Gaps 0:
	Query Match	Score 83.5:	DB 6:	Length 610;	
	Best Local Similarity	53.9%;	Pred. No. 0.00014:		
	Matches 15: Conservative	50.0%;	Mismatches 9:	Indels 0:	Gaps 0:



Search completed: July 3, 2002, 12:40:22  
Job time: 361 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:35:22 : Search time 28.2 Seconds  
(without alignments)  
92.000 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155

Sequence: 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-71.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142	91.6	247	1 LNRBPS	pulmonary surfacta
2	140	90.3	248	2 A48853	pulmonary surfacta
3	137	88.4	248	1 LNHUPS	pulmonary surfacta
4	137	88.4	248	1 LNHUP6	pulmonary surfacta
5	137	88.4	248	1 LNHUPL	pulmonary surfacta
6	136	87.7	248	1 LNRTPS	pulmonary surfacta
7	134	86.5	248	2 I51921	pulmonary surfacta
8	133	85.8	248	1 LNDGFS	pulmonary surfacta
9	108	69.7	375	1 A45225	pulmonary surfacta
10	103	66.5	301	2 A53570	collectin-43 - bov
11	103	66.5	374	1 A42046	surfactant protein
12	100	64.5	369	2 S33603	surfactant protein
13	95	61.3	371	1 JN0450	conglutinin precu
14	95	61.3	371	2 I45878	conglutinin - bovi
15	87	56.1	244	1 LNRTRC	mannose-binding le
16	83	53.5	244	1 LNM5MC	mannose-binding le
17	81	52.3	248	1 LNHUMC	mannose-binding le
18	80.5	51.9	485	2 S36772	E-selectin - bovin
19	79.5	51.3	482	2 JC5092	E-selectin - pig
20	79.5	51.3	830	2 A30359	P-selectin precurs
21	79	51.0	311	1 LNHU2A	asialoglycoprotein
22	77	49.7	238	1 LNRTHA	mannose-binding le
23	72.5	46.8	376	2 JC4892	L-selectin precurs
24	72.5	46.8	551	2 I46709	endothelial leukoc
25	72.5	46.8	610	2 A35046	E-selectin precurs
26	72.5	46.8	612	2 B42755	E-selectin precurs
27	72	46.5	301	2 S13165	asialoglycoprotein
28	71	45.8	239	1 LNM5MA	mannose-binding le
29	70.5	45.5	370	2 S22124	L-selectin precurs

#### ALIGNMENTS

RESULT 1

LNRBPS

pulmonary surfactant protein A precursor - rabbit

N;Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999

C;Accession: A29931

R;Bogdaram, V.; Qing, K.; Mendelson, C.R.

J. Biol. Chem. 263, 2939-2947, 1988

A;Title: The major apoprotein of rabbit pulmonary surfactant. Elucidation of primary

A;Reference number: A29931; MUID:88139348

A;Accession: A29931

A;Molecule type: mRNA

A;Residues: 1-247 <BOG>

A;Cross-references: GB:J03542; NID:g165705; PIDN:AAA31465.1; PID:g165706

A;Note: 12-Pro was also found

A;Note: Two species of mRNA, which appear to be transcribed from a single gene, could

A;Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lower

C;Comment: This protein is a stialoglycoprotein synthesized by alveolar type II cells.

C;Superfamily: mannose-binding lectin; C-type lectin homology

C;Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; gl

F;1-15/Domain: signal sequence #status predicted <SIG>

F;16-247/Product: pulmonary surfactant protein A #status predicted <MAT>

F;27-99/Region: collagen-like

F;126-245/Domain: C-type lectin homology <LCH>

F;16/Modified site: acetylated amino end (Ser) (in mature form) #status predicted

F;206/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 91.68; Score 142; DB 1; Length 247;

Best Local Similarity 88.94; Pred. No. 3.5e-13;

Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27

Db 221 EKCVEMTDGKWNDRNCLOSLRAICEF 247

RESULT 2

A48853

pulmonary surfactant-associated protein SP-A - mouse

C;Species: Mus musculus (house mouse)

C;Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Aug-1999

C;Accession: A48853

R;Korfhagen, T.R.; Bruno, M.D.; Glasser, S.W.; Cirraolo, P.J.; Whitsett, J.A.; Lattier

Am. J. Physiol. 263, L546-L554, 1992

A;Title: Murine pulmonary surfactant SP-A gene: cloning, sequence, and transcriptiona

A;Reference number: A48853; MUID:93072386

A;Contents: DBA/2J

A;Accession: A48853

```

A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-248 <KOR>
A:Cross-references: GB:S48768; NID:g260452; PIDN:AA2474.1; PID:g260453
A>Note: sequence extracted from NCBI backbone (NCBIN:118740, NCBIPI:118741)
C:Superfamily: mannose-binding lectin; C-type lectin homology
F:127-246/Domain: C-type lectin homology <LCH>

Query Match      90.3%; Score 140; DB 2; Length 248;
Best Local Similarity 88.9%; Pred. No. 6.9e-13;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRNCLOSLRAICEF 27
      ||||| ||||| ||||| ||||| |||||
Db 222 EKCVMYTDGKNDRNCLOSLRAICEF 248

RESULT 3
LNHUPS
pulmonary surfactant protein A precursor (genomic clone) - human
N:Alternate names: alveolar proteinosis protein; pulmonary surfactant 32K apoprotein; pu
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
R:White, R.T.; Damm, D.; Miller, J.; Spratt, K.; Schilling, J.; Hawgood, S.; Benson, B.;
Nature 317, 361-363, 1985
A:Title: Isolation and characterization of the human pulmonary surfactant apoprotein gen
A:Reference number: A24622; MUID:86014366
A:Accession: A24622
A:Molecule type: DNA
A:Residues: 1-248 <WHI>
A:Cross-references: GB:M30838; NID:g190564; PIDN:AAA36510.1; PID:g190565
A>Note: the sequence in GenBank entry HUMPSAP, release 109.0, (PID:g190565) has the cod
A>Note: four nucleotide differences, producing amino acid differences at positions 45, 5
R:Hagood, H.P.; White, R.T.; Schilling, J.; Lau, K.; Benson, B.J.; Golden, J.; Hawgood
Am. J. Physiol. 257, L421-L429, 1989
A:Title: Studies of the structure of lung surfactant protein SP-A.
A:Reference number: A43628; MUID:90119861
A:Accession: A43628
A:Molecule type: protein
A:Residues: 143-150;220-240;243-248 <HAA>
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t
C:Comment: This protein is a sialoglycoprotein synthesized by alveolar type II cells. It
pendent on the presence of calcium ions.
C:Genetics:
A:Gene: GDB:SFTPL1; SP-A; SP-A1
A:Cross-references: GDB:119593; OMIM:178630
A:Map position: 10q22-10q23
A:Introns: 58/1; 98/1; 124/1
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: alveolar proteinosis; calcium; gaseous exchange; glycoprotein; hydroxylysine
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-248/Product: pulmonary surfactant protein A #status predicted <MAT>
F:28-100/Domain: collagenous #status predicted <COL>
F:127-246/Domain: C-type lectin homology <LCH>
F:26/Disulfide bonds: interchain #status experimental
F:30,33,36,42,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status predict
F:51.88/Modified site: 5-hydroxylysine (Lys) #status predicted
F:155-246,224-238/Disulfide bonds: #status experimental
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      88.4%; Score 137; DB 1; Length 248;
Best Local Similarity 85.2%; Pred. No. 1.9e-12;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRNCLOSLRAICEF 27
      ||||| ||||| ||||| ||||| |||||
Db 222 EKCVMYTDGKNDRNCLOSLRAICEF 248

RESULT 4
LNHUPS
pulmonary surfactant protein A precursor (clone 6A) - human
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associat
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: A25720
R:Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.; Recny, M.; Sultzman,
J. Biol. Chem. 261, 9029-9033, 1986
A:Title: Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfa
A:Reference number: A25720; MUID:86250832
A:Accession: A25720
A:Molecule type: mRNA
A:Residues: 1-248 <FLO>
A:Cross-references: GB:M13686; NID:g190669; PIDN:AAA60211.1; PID:g190670
A>Note: part of the sequence was confirmed by protein sequencing
A>Note: the amino end of the mature protein, which was not identified, is partially a
A>Note: clones corresponding to two different proteins were sequenced. Cotranslationa
C:Genetics:
A:Gene: GDB:SFTPL1; SP-A; SP-A1
A:Cross-references: GDB:119593; OMIM:178630
A:Map position: 10q22-10q23
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; gl
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-248/Product: pulmonary surfactant protein A #status predicted <MAT>
F:127-246/Domain: C-type lectin homology <LCH>
F:21/Modified site: acetylated amino end (Glu) (in mature form) #status predicted
F:30,33,36,42,54,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status p
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match      88.4%; Score 137; DB 1; Length 248;
Best Local Similarity 85.2%; Pred. No. 1.9e-12;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKNDRNCLOSLRAICEF 27
      ||||| ||||| ||||| ||||| |||||
Db 222 EKCVMYTDGKNDRNCLOSLRAICEF 248

RESULT 5
LNHUP1
pulmonary surfactant protein A precursor (clone 1A) - human
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associat
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: B25720
R:Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.; Recny, M.; Sultzman,
J. Biol. Chem. 261, 9029-9033, 1986
A:Title: Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfa
A:Reference number: A25720; MUID:86250832
A:Accession: B25720
A:Molecule type: mRNA
A:Residues: 1-248 <FLO>
A:Cross-references: GB:K03475
A>Note: part of the sequence was confirmed by protein sequencing
A>Note: the amino end of the mature protein, which was not identified, is partially a
A>Note: clones corresponding to two different proteins were sequenced. Cotranslationa
C:Genetics:
A:Gene: GDB:SFTPL1; SP-A; SP-A1
A:Cross-references: GDB:119593; OMIM:178630
A:Map position: 10q22-10q23
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; gl
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-248/Product: pulmonary surfactant protein A #status predicted <MAT>
F:127-246/Domain: C-type lectin homology <LCH>
F:21/Modified site: acetylated amino end (Glu) (in mature form) #status predicted
F:30,33,36,42,54,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status p
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Best Local Similarity 85.2%; Pred. No. 1.9e-12;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
      :|||||:|||||:|||||:|||||:
Db 222 EQCVMYTDGWNDRNCILYSLTICDF 248

RESULT 6
LNTPS
pulmonary surfactant protein A precursor - rat
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: A29299; J50034; S23183
R:Sano, K.; Fisher, J.; Mason, R.J.; Kuroki, Y.; Schilling, J.; Benson, B.; Voelker, D.
Biochim. Biophys. Res. Commun. 144, 367-374, 1987
A:Title: Isolation and sequence of a cDNA clone for the rat pulmonary surfactant-associated
A:Contents: 0.9 kb cDNA
A:Reference number: A29299; MUID:87213191
A:Accession: A29299
A:Molecule type: mRNA
A:Residues: 1-248 <SAP>
A>Note: part of the sequence, including the amino end of the mature protein, was confirmed
R:Fisher, J.H.; Emrie, P.A.; Shannon, J.; Sano, K.; Hattler, B.; Mason, R.J.
Biochim. Biophys. Acta 950, 338-345, 1988
A:Title: Rat pulmonary surfactant protein A is expressed as two differently sized mRNA s
A:Reference number: J50034; MUID:89000785
A:Contents: 1.6 kb cDNA
A:Accession: J50034
A:Molecule type: mRNA
A:Residues: 1-248 <FIS>
R:Lacaze-Masmonville, T.; Fraslon, C.; Bourbon, J.; Raymondjean, M.; Kahn, A.
Eur. J. Biochem. 206, 613-623, 1992
A:Title: Characterization of the rat pulmonary surfactant protein A promoter.
A:Reference number: S23183; MUID:92298987
A:Accession: S23183
A:Molecule type: DNA
A:Residues: 1-32 <LAC>
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers the
pendent on the presence of calcium ions.
C:Comment: Two species of mRNA, which probably are transcribed from a single gene, have
C:Comment: Size heterogeneity of these proteins arises from posttranslational modification
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: alveolar proteinosis; calcium; gaseous exchange; glycoprotein; hydroxyprolin
F:1-20/Domain: signal sequence #status predicted <Sig>
F:21-248/Product: pulmonary surfactant protein A #status experimental <MAT>
F:37-103/Region: collagen-like
F:127-246/Domain: C-type lectin homology <LCH>
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.7%; Score 136; DB 1; Length 248;
Best Local Similarity 85.2%; Pred. No. 2.6e-12;
Matches 23; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
      :|||||:|||||:|||||:|||||:
Db 222 EQCVMYTDGWNDRNCILYSLTICDF 248

RESULT 7
LNTPS
pulmonary surfactant-associated protein A1 - human
N:Alternate names: SP-A1
C:Species: Homo sapiens (man)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Aug-1999
C:Accession: I51921
R:Katyal, S.L.; Singh, G.; Locker, J.
Am. J. Respir. Cell Mol. Biol. 6, 446-452, 1992
A:Title: Characterization of a second human pulmonary surfactant-associated protein s
A:Reference number: I51921; MUID:92198680
A:Accession: I51921
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-248 <RES>
A:Cross-references: GB:M69519; NID:g338048; PIDN:AAA60319.1; PID:g338049
C:Genetics:
A:Gene: GDB:SFTPA1; SFTPL1; SP-A; SP-A1
A:Cross-references: GDB:I19593; OMIM:178630
A:Map position: 10q22-10q23
A:Introns: 58/1; 98/1; 124/1
C:Superfamily: mannose-binding lectin; C-type lectin homology
F:127-246/Domain: C-type lectin homology <LCH>

Query Match 86.5%; Score 134; DB 2; Length 248;
Best Local Similarity 81.5%; Pred. No. 5.1e-12;
Matches 22; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
      :|||||:|||||:|||||:|||||:
Db 222 EQCVMYTDGWNDRNCILYSLTICDF 248

RESULT 8
LNTPS
pulmonary surfactant protein A precursor - dog
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C:Species: Canis lupus familiaris (dog)
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: A52296; A61227; A60142
R:Benson, B.; Hawgood, S.; Schilling, J.; Clements, J.; Damm, D.; Cordell, B.; White,
Proc. Natl. Acad. Sci. U.S.A. 82, 6379-6383, 1985
A:Title: Structure of canine pulmonary surfactant apoprotein: cDNA and complete amino
A:Reference number: A52296; MUID:86016705
A:Accession: A52296
A:Molecule type: mRNA
A:Residues: 1-248 <BEN>
A>Note: the authors translated the codon TGC for residue 60 as Pro
A>Note: part of the sequence, including the amino end of the mature protein, was conf
R:Liau, D.F.; Ryan, S.F.
Chem. Phys. Lipids 59, 29-38, 1991
A:Title: Purification of surfactant protein A from dog lung by reconstitution with su
A:Reference number: A61227; MUID:92163993
A:Accession: A61227
A:Molecule type: protein
A:Residues: 18-32 <LIA>
R:Ross, G.F.; Meuth, J.; Ohning, B.; Kim, Y.; Whitsett, J.A.
Biochim. Biophys. Acta 870, 267-278, 1986
A:Title: Purification of canine surfactant-associated glycoproteins A. Identification
A:Reference number: A60142; MUID:86159848
A:Accession: A60142
A:Molecule type: protein
A:Residues: 24-34;95-101,'X',103-108 <ROS>
R:Patthy, L.
Nature 325, 490, 1987
A:Reference number: A93388; MUID:87115834
A:Contents: annotation; animal lectin domain homology
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lower
pendent on the presence of calcium ions.
C:Superfamily: mannose-binding lectin; C-type lectin homology
C:Keywords: alveolar proteinosis; calcium; disulfide bond; gaseous exchange; hydroxyp
F:1-17/Domain: signal sequence #status predicted <Sig>
F:18-248/Product: pulmonary surfactant protein A #status experimental <MPT>
F:28-102/Region: collagen-like
F:127-246/Domain: C-type lectin homology <LCH>
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:30/Modified site: 4-hydroxyproline (Pro) #status experimental

```







RESULT 14  
145878  
conglutinin - Bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 19-Dec-1997 #sequence\_revision 19-Dec-1997 #text\_change 20-Aug-1999  
C:Accession: I45878  
R:Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry  
Gene 141, 277-281, 1994  
A:Title: Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization c  
A:Reference number: 145878; MUID:94215917  
A:Accession: I45878  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-371 <L10>  
A:Cross-references: GB:L18871; NID:g495012; PIDN:AAA20126.1; PID:g495013  
C:Superfamily: Pulmonary surfactant protein D; C-type lectin homology  
F:248-369/Domain: C-type lectin homology <LCH>

Query Match 61.3%; Score 95; DB 2; Length 371;  
Best Local Similarity 59.3%; Pred. No. 3.7e-06;  
Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAICEF 27  
Db 345 ENCVELFPDGKWNVDPCSKQLLVICF 371  
|||||:|||||:|:|:|

RESULT 15  
LNRTWC  
mannose-binding lectin C precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 22-Jun-1999  
C:Accession: A24791; A38322; JX0201; A26798  
R:Drickamer, K.; Dordal, M.S.; Reynolds, L.  
J. Biol. Chem. 261, 6878-6887, 1986  
A:Title: Mannose-binding proteins isolated from rat liver contain carbohydrate-recogniti  
rotein.  
A:Reference number: A24791; MUID:86196130  
A:Accession: A24791  
A:Molecule type: mRNA  
A:Residues: 1-244 <DRI>  
A:Cross-references: GB:M14103  
A:Note: Part of the sequence, including the amino end of the mature protein, was confir  
R:Childs, R.A.; Felzi, T.; Yuen, C.T.; Drickamer, K.; Quesenberry, M.S.  
J. Biol. Chem. 265, 20770-20777, 1990  
A:Title: Differential recognition of core and terminal portions of oligosaccharide ligand  
A:Reference number: A38322; MUID:91065871  
A:Accession: A38322  
A:Molecule type: mRNA  
A:Residues: 86, 'EL', 89-97 <CHI>  
R:Wada, M.; Itoh, N.; Ohta, M.; Kawasaki, T.  
J. Biochem. 111, 66-73, 1992  
A:Title: Characterization of rat liver mannan-binding protein gene.  
A:Reference number: JX0201; MUID:92299655  
A:Accession: JX0201  
A:Molecule type: DNA  
A:Residues: 1-244 <WAD>  
A:Experimental source: liver  
A:Note: the authors translated the codon CCA for residue 43 as Phe  
R:Oka, S.; Itoh, N.; Kawasaki, T.; Yamashina, I.  
J. Biochem. 101, 135-144, 1987  
A:Title: Primary structure of rat liver mannan-binding protein deduced from its cDNA seq  
A:Reference number: A26798; MUID:87194686  
A:Accession: A26798  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-244 <OKA>  
A:Cross-references: GB:X05023; NID:g56634; PIDN:CAA28697.1; PID:g56635  
C:Comment: The molecule contains six identical chains, occurring as disulfide-bonded dim  
C:Genetics: Mannan-binding proteins are calcium ion-dependent and are specific for mannos  
A:Introns: 59/1; 98/1; 121/1

C:Superfamily: mannose-binding lectin; C-type lectin homology  
C:Keywords: calcium; endoplasmic reticulum; Golgi apparatus; homohexamers; hydroxyprol  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-244/Product: mannose-binding lectin C #status experimental <MAT>  
F:38-94/Region: collagen-like  
F:124-240/Domain: C-type lectin homology <LCH>  
F:29.34/Disulfide bonds: interchain #status predicted  
F:69/Modified site: 4-hydroxyproline (Pro) #status experimental

Query Match 56.1%; Score 87; DB 1; Length 244;  
Best Local Similarity 55.6%; Pred. No. 3.6e-05;  
Matches 15; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAICEF 27  
Db 216 ENCVELFPDGKWNVDPCSKQLLVICF 242  
|||||:|||||:|:|:|

Search completed: July 3, 2002, 12:35:22  
Job time: 191 sec

GenCore version 4.5  
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:40:43 ; Search time 15.79 Seconds  
(without alignments)  
66.208 Million cell updates/sec

Title: US-09-600-932-3

Perfect score: 155

Sequence: 1 EKCVEMTDGKWNDRCLQSLRAICEF 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	142	91.6	247	1 PSPA_RABIT	P12842 oryctolagus
2	140	90.3	248	1 PSPA_MOUSE	P35242 mus musculus
3	139	89.7	249	1 PSPA_PIG	P49874 sus scrofa
4	137	88.4	248	1 PSPA_HUMAN	P07714 homo sapien
5	136	87.7	248	1 PSPA_RAT	P08427 rattus norv
6	133	85.8	248	1 PSPA_CANFA	P06908 canis fami
7	132	85.2	247	1 PSPA_CAVPO	P50403 cavia porce
8	108	69.7	375	1 PSPD_HUMAN	P35247 homo sapien
9	103	66.5	301	1 CL43_BOVIN	P42916 bos taurus
10	103	66.5	374	1 PSPD_MOUSE	P50404 mus musculus
11	103	66.5	374	1 PSPD_RAT	P35248 rattus norv
12	100	64.5	369	1 PSPD_BOVIN	P35246 bos taurus
13	95	61.3	371	1 CONG_BOVIN	P23805 bos taurus
14	87	56.1	244	1 MABC_RAT	P08661 rattus norv
15	83	53.5	244	1 MABC_MOUSE	P41117 mus musculus
16	81	52.3	248	1 MABC_HUMAN	P11226 homo sapien
17	81	52.3	249	1 MABC_BOVIN	P02659 bos taurus
18	80.5	51.9	485	1 LEM2_BOVIN	P98107 bos taurus
19	79.5	51.3	484	1 LEM2_PIG	P98110 sus scrofa
20	79.5	51.3	830	1 LEM3_HUMAN	P16109 homo sapien
21	79	51.0	311	1 LEC1_HUMAN	P07307 homo sapien
22	77.5	50.0	611	1 LEM2_CANFA	P33730 canis fami
23	77	49.7	238	1 MABA_RAT	P19999 rattus norv
24	72.5	46.8	551	1 LEM2_RABIT	P27113 oryctolagus
25	72.5	46.8	610	1 LEM2_HUMAN	P16581 homo sapien
26	72.5	46.8	612	1 LEM2_MOUSE	P00690 mus musculus
27	72	46.5	301	1 LEC1_MOUSE	P24721 mus musculus
28	71.5	46.1	549	1 LEM2_RAT	P98105 rattus norv
29	71	45.8	239	1 MABA_MOUSE	P39039 mus musculus
30	70.5	45.5	370	1 LEM1_BOVIN	P98131 bos taurus
31	69.5	44.8	768	1 LEM3_MOUSE	P01102 mus musculus
32	69.5	44.8	768	1 LEM3_RAT	P98106 rattus norv
33	68.5	44.2	372	1 LEM1_MOUSE	P18337 mus musculus

34	68.5	44.2	372	1 LEM1_RAT	P30836 rattus norv
35	67.5	43.5	372	1 LEM1_MACMU	Q95198 macaca mulia
36	67.5	43.5	372	1 LEM1_PAPHA	Q28768 papio hamad
37	67.5	43.5	646	1 LEM3_BOVIN	P42201 bos taurus
38	67	43.2	301	1 LEC1_RAT	P08290 rattus norv
39	66.5	42.9	372	1 LEM1_HUMAN	P14151 homo sapien
40	66.5	42.9	372	1 LEM1_PANTR	Q95237 pan troglod
41	66.5	42.9	372	1 LEM1_PONPY	Q95235 pongo pygma
42	65.5	42.3	769	1 LEM3_SHEEP	P98109 ovis aries
43	65	41.9	147	1 PL1A_TRIFL	P21755 trimeresuru
44	65	41.9	197	1 CLF1_HUMAN	O75596 homo sapien
45	65	41.9	283	1 LEM3_MOUSE	P34927 mus musculus

#### ALIGNMENTS

```

RESULT 1
PSPA_RABIT
ID PSPA_RABIT STANDARD; PRT; 247 AA.
AC P12842;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPAP).
GN SFTPA1 OR SFTPA OR SFTPL
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88139348; PubMed=2830270;
RA Boggaram V., Qing K., Mendelson C.R.;
RT "The major apoprotein of rabbit pulmonary surfactant. Elucidation of
RT primary sequence and cyclic AMP and developmental regulation.";
RL J. Biol. Chem. 263:2939-2947(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;
RX MEDLINE=92312742; PubMed=1616051;
RA Chen Q., Boggaram V., Mendelson C.R.;
RT "Rabbit lung surfactant protein A gene: identification of a lung-
RT specific DNase I hypersensitive site.";
RL Am. J. Physiol. 262:L662-L671(1992).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSPAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J03542; AAA31465.1; -
CC EMBL; L19187; AAA31468.1; -
CC PIR; A29931; LNRBPS.
CC HSP; P35247; I808.
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR001304; lectin_c.

```

```
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE: PS00041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 247 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT DOMAIN 27 99 A.
FT DOMAIN 152 247 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 154 245 BY SIMILARITY.
FT DISULFID 223 237 BY SIMILARITY.
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (PROBABLE).
FT VARIANT 12 12 S -> P.
FT CONFLICT 57 60 GPMG -> APWA (IN REF. 2).
SQ SEQUENCE 247 AA; 26071 MW; 289634054C8C84 CRC64;

Query Match 91.6%; Score 142; DB 1; Length 247;
Best Local Similarity 88.9%; Pred. No. 1.le-13;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
|||||
DB 221 EKCVMYTDGKWNDRNCLOSLAICEF 247

RESULT 2
PSPA_MOUSE
ID PSPA_MOUSE STANDARD; PRT; 248 AA.
AC P35242;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPAP).
GN SFTPA1 OR SFTPA OR SFTP1 OR SFTP-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DNA/2J;
RX MEDLINE=93072386; PubMed=1443158;
RA Korfhagen T.R., Bruno M.D., Glasser S.W., Ciraoilo P.J., Whitsett J.A.,
RA Lattier D.L., Wikenheiser K.A., Clark J.C.;
RT "Murine pulmonary surfactant SP-A gene: cloning, sequence, and
transcriptional activity.";
RL Am. J. Physiol. 263:L546-L554(1992).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S48768; AAB24274.1; -.
```

```
DR PIR: A48853; A48853.
DR HSSP: P35247; 1B08.
DR MGD; MG1:109518; Sftpa.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE: PS00041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 248 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT DOMAIN 28 100 COLLAGEN-LIKE.
FT DOMAIN 153 248 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 155 246 BY SIMILARITY.
FT DISULFID 224 238 BY SIMILARITY.
FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (PROBABLE).
SQ SEQUENCE 248 AA; 26157 MW; 6688BF070E3EB9AE CRC64;

Query Match 90.3%; Score 140; DB 1; Length 248;
Best Local Similarity 88.9%; Pred. No. 2.le-13;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSLAICEF 27
|||||
DB 222 EKCVMYTDGKWNDRNCLOSLAICEF 248

RESULT 3
PSPA_PIG
ID PSPA_PIG STANDARD; PRT; 249 AA.
AC P49874;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPAP).
GN SFTPA1 OR SFTPA OR SFTP1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Adamou J.E., Elshourbagy N.A.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C). (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L41350; AAA88403.1; -.
```

```

DR HSP: P19999; LYTT.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS00411; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat
FT SIGNAL 1 20
FT CHAIN 21 249
FT DOMAIN 28 100
FT DOMAIN 153 249
FT DISULFID 155 247
FT DISULFID 225 239
FT CARBOHYD 208 208
SQ SEQUENCE 249 AA; 26702 MW; 3C4E05AD07F2A7CD CRC64;

Query Match 89.7%; Score 139; DB 1; Length 249;
Best Local Similarity 88.9%; Pred. No. 3e-13; Indels 0; Gaps 0;
Matches 24; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKVEMTDGKNDNRNCQLSLAICEF 27
      |||||
DB 223 EKVEMTDGQWDRNCQYRLAICEF 249

RESULT 4
ID PSPA_HUMAN STANDARD; PRT; 248 AA.
AC P07714;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)
DE (PSPAP) (Alveolar proteinosis protein) (35 kDa pulmonary surfactant-
DE associated protein).
GN SPTPA1 OR SPTPA OR SPTP1 OR PSPAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86250832; PubMed=3755136;
RA Floros J., Steinbrink R., Jacobs K., Phelps D., Kriz R., Recny M.,
RA Sultman L., Jones S., Taesch H.W., Frank H.A., Fritsch E.F.;
RT "Isolation and characterization of cDNA clones for the 35-kDa
RT pulmonary surfactant-associated protein.";
RL J. Biol. Chem. 261:9029-9033(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86014366; PubMed=2995821;
RA White R.T., Damm D., Miller J., Spratt K., Schilling J., Hawgood S.,
RA Benson B., Cordell B.;
RT "Isolation and characterization of the human pulmonary surfactant
RT apoprotein gene";
RL Nature 317:361-363(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92198680; PubMed=1372511;
RA Katyal S.L., Singh G., Locker J.L.;
RT "Characterization of a second human pulmonary surfactant-associated
RT protein SP-A gene.";
RL Am. J. Respir. Cell Mol. Biol. 6:446-452(1992).
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSPAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.

```

```

CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M13686; AAA60211.1; -
CC EMBL; K03475; AAA36520.1; -
CC EMBL; M30838; AAA36510.1; -
CC EMBL; M68519; AAA60319.1; -
CC PIR; A24622; LNHUP5.
CC PIR; A25720; LNHUP6.
CC PIR; B25720; LNHUP1.
CC HSP; P19999; LYTT.
CC MIM; 178630; -
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001304; lectin_c.
CC Pfam; PF00059; lectin_c; 1.
CC SMART; SM00034; CLECT; 1.
CC PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
CC PROSITE; PS00411; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lectin; Collagen; Repeat; Polymorphism.
FT SIGNAL 1 20
FT CHAIN 21 248
FT DOMAIN 28 100
FT DOMAIN 153 248
FT DISULFID 155 246
FT DISULFID 224 238
FT CARBOHYD 207 207
FT VARIANT 9 9
FT VARIANT 50 50
FT VARIANT 66 66
FT VARIANT 73 73
FT VARIANT 81 81
FT VARIANT 85 85
FT VARIANT 219 219
FT VARIANT 223 223
FT VARIANT 19 19
FT VARIANT 45 45
FT VARIANT 54 54
FT VARIANT 91 91
FT VARIANT 100 100
FT VARIANT 247 247
FT SEQUENCE 248 AA; 26214 MW; 6A9F0C3488BF3633 CRC64;

Query Match 88.4%; Score 137; DB 1; Length 248;
Best Local Similarity 85.2%; Pred. No. 5.8e-13;
Matches 23; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKVEMTDGKNDNRNCQLSLAICEF 27
      |||||
DB 222 EKVEMTDGQWDRNCQLSLAICEF 248

```

```
RESULT 5
PSPA_RAT
ID PSPA_RAT STANDARD; PRT; 248 AA.
AC P08427;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A) (PSPAP).
GN SFTPA1 OR SFTPA OR SFTP1 OR SFTP-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=89000785; PubMed=2901856;
RA Fisher J.H., Emrie P.A., Shannon J., Sano K., Hattler B., Mason R.J.;
RT "Rat pulmonary surfactant protein A is expressed as two differently
RT sized mRNA species which arise from differential polyadenylation of
RT one transcript.";
RL Biochim. Biophys. Acta 950:338-345(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87213191; PubMed=3579914;
RA Sano K., Fisher J.H., Mason R.J., Kuroki Y., Schilling J., Benson B.,
RA Voelker D.;
RT "Isolation and sequence of a cDNA clone for the rat pulmonary
RT surfactant-associated protein (PSP-A).";
RL Biochem. Biophys. Res. Commun. 144:367-374(1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Lung;
RX MEDLINE=96088665; PubMed=7491978;
RA Smith C.I., Rosenberg E., Reisher S.R., Li F., Kefalides P.,
RA Fisher A.B., Feinstein S.I.;
RT "Sequence of rat surfactant protein A gene and functional mapping of
RT its upstream region.";
RL Am. J. Physiol. 269:L603-L612(1995).
CC -1- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M33201; AAA41973.1; -
DR EMBL; X13176; CAA31574.1; -
DR EMBL; X13177; CAA31574.1; -
DR EMBL; M15754; AAA41972.1; ALT_INIT.
DR EMBL; U43092; AAA85516.1; -
DR PIR; A29299; LNRTPS.
DR HSSP; P35247; 1B08.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.

DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS00411; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
FT SIGNAL; 1 20 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT CHAIN; 21 248 A.
FT DOMAIN; 28 100 COLLAGEN-LIKE.
FT DOMAIN; 153 248 C-TYPE LECTIN (SHORT FORM).
FT DISULFID; 155 246 BY SIMILARITY.
FT DISULFID; 224 238 BY SIMILARITY.
FT CARBOHYD; 21 21 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD; 207 207 N-LINKED (GLCNAC... ) (PROBABLE).
FT CONFLICT; 78 78 A -> G (IN REF. 1).
FT CONFLICT; 84 84 E -> G (IN REF. 1).
FT CONFLICT; 139 139 MISSING (IN REF. 2).
FT CONFLICT; 156 156 T -> TF (IN REF. 2).
FT CONFLICT; 180 180 K -> N (IN REF. 1; CAA31574).
SQ SEQUENCE 248 AA; 26288 MW; CAA02030095682A5 CRC64;

Query Match 87.7%; Score 136; DB 1; Length 248;
Best Local Similarity 85.2%; Pred. No. 8.1e-13;
Matches 23; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EKCVMYTDGKWNDRNCLOSRLAICEF 27
|||||
DB 222 EKCVMYTDGWNDRGCLQYRLAVCEF 248
|||||

RESULT 6
PSPA_CANFA
ID PSPA_CANFA STANDARD; PRT; 248 AA.
AC P06908;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A) (PSPAP).
DE (PSPAP).
GN SFTPA1 OR SFTPA OR SFTP1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND SIGNAL SEQUENCE CLEAVAGE SITE.
RX MEDLINE=86016705; PubMed=3863100;
RA Benson B., Hawgood S., Schilling J., Clements J., Damm D., Cordell B.,
RA White R.T.;
RT "Structure of canine pulmonary surfactant apoprotein: cDNA and
RT complete amino acid sequence.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6379-6383(1985).
RN [2]
RP C-TYPE LECTIN DOMAIN.
RX MEDLINE=87115834; PubMed=3808053;
RA Patthy L.;
RT "Is lung surfactant protein a lectin-collagen hybrid?";
RL Nature 325:490-490(1987).
CC -1- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

RESULT	8
PSPD_HUMAN	
ID	PSPD_HUMAN
AC	STANDARD;
PR	375 AA.
DT	P35247;
DT	01-FEB-1994 (Rel. 28, Created)
DT	01-FEB-1994 (Rel. 28, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN	SFTPD OR SFTPA OR PSPD.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;	
[1]	
RP	SEQUENCE FROM N.A.
RC	TISSUE=Placenta;
RA	MEDLINE=931515122; PubMed=8428971;
RX	Crouch E., Rust K., Veille R., Denis-Keller H., Grosso L.;
RT	"Genomic organization of human surfactant protein D (SP-D) is
RT	encoded on chromosome 10q22.2-23.1.";
RL	J. Biol. Chem. 268:2976-2983(1993).
[2]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 214-243.
RC	TISSUE=Lung, and Amniotic fluid;
RX	MEDLINE=92322003; PubMed=1339284;
RA	Lu J., Willis A.C., Reid K.B.M.;
RT	"Purification, characterization and cDNA cloning of human lung
RT	surfactant protein D.";
RL	Biochem. J. 284:795-802(1992).
[3]	
RP	SEQUENCE OF 60-375 FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Lung;



SQ SEQUENCE 375 AA; 37702 MW; 2986B2699FC01A6A CRC64;  
  
 Query Match 69.7%; Score 108; DB 1; Length 375;  
 Best Local Similarity 63.0%; Pred No. 1.5e-08;  
 Matches 17; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
  
 QY 1 EKCVMTDQKWNDRNCLOSLRAICEF 27  
 I :  
 Db 349 EDCVEIFTNGKWDNRCKGEKRLVWCEF 375  
  
 RESULT 9  
 CL43\_BOVIN STANDARD; PRT; 301 AA.  
 AC P42916;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Collectin-43 (CL-43).  
 OS Bos taurus (Bovidae).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovioidea;  
 OC Bovidae; Bovinae; Bos.  
 NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Liver.  
 RX MEDLINE=94216283; PubMed=8163480;  
 RA Lim B.-L., Willis A.C., Reid K.B.M., Lu J., Laursen S.B.,  
 RA Jensenius J.C., Holmskov U.;  
 RT "Primary structure of bovine collectin-43 (CL-43). Comparison with  
 conglutinin and lung surfactant protein-D";  
 RL J. Biol. Chem. 269:11820-11824(1994)  
 CC -!- FUNCTION: LECTIN THAT BINDS TO VARIOUS SUGARS: MANNOSE >  
 CC FUCOSE > GLCNAC > GLUCOSE = MALTOSE > GALACTOSE > LACTOSE >  
 CC GALNA. COULD PLAY A ROLE IN IMMUNE DEFENSE.  
 CC -!- SUBUNIT: OLGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL; X75912; CAA53511.1; ALT\_SEQ.  
 DR HSSP; P35247; 1B08.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001304; lectin\_c.  
 DR Pfam; PF01391; Collagen; 2.  
 DR Pfam; PR00059; lectin\_c; 1.  
 DR SMART; SM00034; CLECT; 1.  
 DR PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
 DR PROSITE; PS00041; C-TYPE LECTIN\_2; 1.  
 KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Collagen;  
 KW Repeat; Calcium.  
 FT DOMAIN 29 142 COLLAGEN-LIKE.  
 FT DOMAIN 202 301 C-TYPE LECTIN (SHORT FORM).  
 FT DISULFID 204 299 BY SIMILARITY.  
 FT DISULFID 277 291 BY SIMILARITY.  
 SQ SEQUENCE 301 AA; 31362 MW; 0385C10B8424CD76 CRC64;  
  
 Query Match 66.5%; Score 103; DB 1; Length 301;  
 Best Local Similarity 59.3%; Pred No. 6.6e-08;  
 Matches 16; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
  
 QY 1 EKCVMTDQKWNDRNCLOSLRAICEF 27

Query Match 66.5%; Score 103; DB 1; Length 374;  
Best Local Similarity 59.3%; Pred.No.8.1e-08;  
Matches 16; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

```

ET DOMAIN 45 221 COLLAGEN-LIKE.
FT DOMAIN 222 253 COILED COIL (POTENTIAL).
FT DOMAIN 278 374 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 280 372 BY SIMILARITY.
FT DISULFID 350 364 BY SIMILARITY.
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .).
FT MOD_RES 77 77 HYDROXYLATION.
FT MOD_RES 86 86 HYDROXYLATION.
FT MOD_RES 95 95 HYDROXYLATION.
FT MOD_RES 98 98 HYDROXYLATION.
FT MOD_RES 170 170 HYDROXYLATION.
FT MOD_RES 176 176 HYDROXYLATION.
FT CONFLICT 89 89 N -> E (IN REF. 2).
FT CONFLICT 164 164 K -> C (IN REF. 2).
SQ SEQUENCE 374 AA; 37561 MW; DB2BB5E399DB4A3C CRC64;

Query Match 66.5%; Score 103; DB 1; Length 374;
Best Local Similarity 59.3%; Pred. No. 8.3e-08;
Matches 16; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27
| |||:::||||:|:| |||||
Db 348 ENCVETFGQNDKACGEQRLVICEF 374

RESULT 12
PSPD_BOVIN
ID PSPD_BOVIN STANDARD; PRT; 369 AA.
AC P35246;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN SFTPD OR SFTPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
RN [1]
RC SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
RX TISSUE=Lung;
RM MEDLINE=93170856; PubMed=8436402;
RA Lim B.L., Lu J., Reid K.B.M.;
RT "Structural similarity between bovine conglutinin and bovine lung surfactant protein D and demonstration of liver as a site of synthesis of conglutinin.";
RL Immunology 78:159-165(1993).
CC -!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL CARBOHYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75911; CAA53510.1;
CC PIR; S33603; S33603.
CC HSSP; P35247; 1B08.

```

```

DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
FT SIGNAL 1 20 Coiled coil.
FT CHAIN 21 369 BY SIMILARITY.
FT FT PULMONARY SURFACTANT-ASSOCIATED PROTEIN D.
FT DOMAIN 46 216 COLLAGEN-LIKE.
FT DOMAIN 217 248 COILED COIL (POTENTIAL).
FT DOMAIN 273 369 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 275 367 BY SIMILARITY.
FT DISULFID 345 359 BY SIMILARITY.
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MOD_RES 78 78 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 87 87 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 96 96 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 99 99 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 165 165 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 171 171 HYDROXYLATION (BY SIMILARITY).
SQ SEQUENCE 369 AA; 37361 MW; 07D88B24E0AEB2E3 CRC64;

Query Match 64.5%; Score 100; DB 1; Length 369;
Best Local Similarity 59.3%; Pred. No. 2.2e-07;
Matches 16; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 EKCVEMTDGKWNDRNCLOSLRAICEF 27
| |||:::||||:|:| |||||
Db 343 ENCVETFGQNDKACGEQRLVICEF 369

RESULT 13
CONG_BOVIN
ID CONG_BOVIN STANDARD; PRT; 371 AA.
AC P23805;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Conglutinin precursor.
GN CGNI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RC SEQUENCE FROM N.A.
RX MEDLINE=93277452; PubMed=7684896;
RA Lu J., Laursen S.B., Thiel S., Jensenius J.C., Reid B.M.;
RT "The cDNA cloning of conglutinin and identification of liver as a primary site of synthesis of conglutinin in members of the Bovidae.";
RL Biochem. J. 292:157-162(1993).
RN [3]
RC SEQUENCE FROM N.A.
RX TISSUE=Liver;
RA MEDLINE=94215917; PubMed=8163202;
RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B., Tauber A.I., Sastry K.N.;
RT "Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization of the BC cDNA reveals strong homology to surfactant protein-D.";
RL Gene 141:277-281(1994).

```

RN SEQUENCE FROM N.A.  
 RP TISSUE-Semen;  
 RX MEDLINE=94267222; PubMed=8207234;  
 RA Liou L.-S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,  
 RA Tauber A.I., Sastry K.N.;  
 RT "Bovine conglutinin gene exon structure reveals its evolutionary  
 RT relationship to surfactant protein-D";  
 RL J. Immunol. 153:173-180(1994).  
 RN [5]  
 RP SEQUENCE OF 21-371.  
 RX MEDLINE=91131556; PubMed=1993651;  
 RA Lee Y.-M., Leiby K.R., Allar J., Paris K., Lerch B., Okarma T.B.;  
 RT "Primary structure of bovine conglutinin, a member of the C-type  
 RT animal lectin family";  
 RL J. Biol. Chem. 266:2715-2723(1991).  
 RN [6]  
 RP PRELIMINARY SEQUENCE OF 21-52.  
 RX MEDLINE=87184551; PubMed=3566740;  
 RA Young N.M., Leon M.A.;  
 RT "The carbohydrate specificity of conglutinin and its homology to  
 RT proteins in the hepatic lectin family";  
 RL Biochem. Biophys. Res. Commun. 143:645-651(1987).  
 CC -!- FUNCTION: CALCIUM-DEPENDENT LECTIN-LIKE PROTEIN WHICH BINDS TO A  
 CC YEAST CELL WALL EXTRACT & IMMUNE COMPLEXES THROUGH THE COMPLEMENT  
 CC COMPONENT (C3BI). IT IS CAPABLE OF BINDING NONREDUCING TERMINAL  
 CC N-ACETYLGLUCOSAMINE, MANNOSE, AND FUCOSE RESIDUES.  
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.  
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D14085; BAA03170.1; -  
 DR EMBL: X71774; CAA50665.1; -  
 DR EMBL: L18871; AAA20126.1; -  
 DR EMBL: U06860; AAB60624.1; -  
 DR EMBL: U06854; AAB60624.1; JOINED.  
 DR EMBL: U06855; AAB60624.1; JOINED.  
 DR EMBL: U06856; AAB60624.1; JOINED.  
 DR EMBL: U06857; AAB60624.1; JOINED.  
 DR EMBL: U06858; AAB60624.1; JOINED.  
 DR PIR: A29416; A29416.  
 DR PIR: A23740; A23740.  
 DR PIR: JN0450; JN0450.  
 DR HSP: P35247; I1B08.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR001304; lectin\_c.  
 DR Pfam: PF01391; Collagen; 3.  
 DR Pfam: PF00059; lectin\_c; 1.  
 DR SMART: SM00034; CLECT; 1.  
 DR PROSITE: PS00615; C-TYPE LECTIN\_1; 1.  
 DR PROSITE: PS50041; C-TYPE LECTIN\_2; 1.  
 KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;  
 KW Collagen; Repeat; Calcium; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 371 CONGLUTININ.  
 FT DOMAIN 45 216 COLLAGEN-LIKE.  
 FT DOMAIN 273 371 C-TYPE LECTIN (SHORT FORM).  
 FT MOD\_RES 63 63 HYDROXYLATION.  
 FT MOD\_RES 87 87 HYDROXYLATION.  
 FT MOD\_RES 99 99 HYDROXYLATION.  
 FT MOD\_RES 135 135 HYDROXYLATION.  
 FT MOD\_RES 141 141 HYDROXYLATION.  
 FT MOD\_RES 159 159 HYDROXYLATION.  
 FT MOD\_RES 162 162 HYDROXYLATION.

FT MOD\_RES 198 198 HYDROXYLATION.  
 FT SITE 201 203 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DISULFID 275 369 BY SIMILARITY.  
 FT DISULFID 347 361 BY SIMILARITY.  
 FT CARBOHYD 337 173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 173 173 R -> H (IN REF. 2 AND 3).  
 FT CONFLICT 210 210 K -> S (IN REF. 5).  
 FT CONFLICT 218 218 V -> A (IN REF. 2).  
 FT CONFLICT 272 272 E -> V (IN REF. 2).  
 SQ SEQUENCE 371 AA; 37994 MW; 867BB41992544B1F CRC64;  
 Query Match 61.3%; Score 95; DB 1; Length 371;  
 Best Local Similarity 59.3%; Pred. No. 1.2e-06;  
 Matches 16; Conservative 3; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 EKCVMTYDGKWNDRNCISRLAICEF 27  
 DQ 345 ENCVETFPDGKWNDRVPCSKOLLVCEF 371  
 RESULT 14  
 MABC.RAT  
 ID MABC.RAT STANDARD; PRT; 244 AA.  
 AC P08661;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)  
 DE (RA-reactive factor P28A subunit) (RARE/P28A).  
 GN MB12.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=92299655; PubMed=1607365;  
 RA Wada M., Itoh N., Ohta M., Kawasaki T.;  
 RT "Characterization of rat liver mannan-binding protein gene.";  
 RL J. Biochem. 111:66-73(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=86196130; PubMed=3009480;  
 RA Drickamer K., Dordal M.S., Reynolds L.;  
 RT "Mannose-binding proteins isolated from rat liver contain  
 RT carbohydrate-recognition domains linked to collagenous tails.  
 RT Complete primary structures and homology with pulmonary surfactant  
 RT apoprotein.";  
 RL J. Biol. Chem. 261:6878-6887(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Liver;  
 RX MEDLINE=87194686; PubMed=3032924;  
 RA Oka S., Itoh N., Kawasaki T., Yamashina I.;  
 RT "Primary structure of rat liver mannan-binding protein deduced from  
 RT its cDNA sequence.";  
 RL J. Biochem. 101:135-144(1987).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 133-244.  
 RX MEDLINE=96132792; PubMed=8557671;  
 RA Ng K.K.-S., Drickamer K., Weis W.I.;  
 RT "Structural analysis of monosaccharide recognition by rat liver  
 RT mannose-binding protein.";  
 RL J. Biol. Chem. 271:663-674(1996).  
 CC -!- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-  
 CC DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS.  
 CC BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF  
 CC THE ANTIBODY.  
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.  
 CC -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR. MOST

CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI  
 CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND  
 CC LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNALE  
 CC SPACE OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE  
 CC PROTEIN.  
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M14103; AAA41554.1; -.  
 CC EMBL; X05023; CAA28687.1; -.  
 CC PIR; A24791; LNRPMC.  
 CC PIR; A26798; A26798.  
 CC PIR; JX0201; JX0201.  
 CC PDB; 1RDI; 08-MAR-96.  
 CC PDB; 1RDJ; 08-MAR-96.  
 CC PDB; 1RDK; 08-MAR-96.  
 CC PDB; 1RDL; 08-MAR-96.  
 CC PDB; 1RDM; 08-MAR-96.  
 CC PDB; 1RDN; 08-MAR-96.  
 CC PDB; 1RDO; 08-MAR-96.  
 CC InterPro; IPR000087; Collagen.  
 CC InterPro; IPR001304; lectin\_c.  
 CC Pfam; PF00059; lectin\_c; 1.  
 CC SMART; SM00034; CLECT; 1.  
 CC PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
 CC PROSITE; PS50041; C-TYPE LECTIN\_2; 1.  
 CC Lectin; Hydroxylation; Liver; Glycoprotein; Mannose-binding; Membrane;  
 CC Calcium; Collagen; Repeat; Signal; 3D-structure.  
 CC SIGNAL 1 18  
 FT CHAIN 19 244 MANNOSE-BINDING PROTEIN C.  
 FT DOMAIN 38 96 COLLAGEN-LIKE.  
 FT DOMAIN 149 242 C-TYPE LECTIN (SHORT FORM).  
 FT MOD\_RES 43 43 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 58 58 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 69 69 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 78 78 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 81 81 HYDROXYLATION (POTENTIAL).  
 FT DISULFID 29 29 INTERCHAIN.  
 FT DISULFID 34 34 INTERCHAIN.  
 FT DISULFID 151 240  
 FT DISULFID 218 232  
 FT CONFLICT 38 39  
 SQ SEQUENCE 244 AA; 26014 MW; F0706E2AA9331531 CRC64;

Query Match 56.1%; Score 87; DB 1; Length 244;  
 Best Local Similarity 55.6%; Pred. No. 1.2e-05;  
 Matches 15; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 EKCVEMYTGKWNDRNCIQSLATCEF 27  
 DB 216 ENCVLLTGKWNDRVPCDSFLVWCF 242

RESULT 15  
 ID MABC\_MOUSE STANDARD; PRT; 244 AA.  
 AC P41317;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)  
 DE (RA-reactive factor P28A subunit) (RAFP/P28A).  
 GN MBL2.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBA/J; TISSUE=Liver;  
 RA MEDLINE=91302823; PubMed=1712818;  
 RX Sastry K., Zahedi K., Lellias J.M., Whitehead A.S., Ezekowitz R.A.;  
 RT "Molecular characterization of the mouse mannose-binding proteins.  
 RL The mannose-binding protein A but not C is an acute phase reactant.";  
 RL J. Immunol. 147:692-697(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALE/C;  
 RA MEDLINE=95284466; PubMed=7766991;  
 RX Sastry R., Wang J.S., Brown D.C., Ezekowitz R.A., Tauber A.I.,  
 RA Sastry K.N.;  
 RT "Characterization of murine mannose-binding protein genes Mbl1 and  
 RL Mbl2 reveals features common to other collectin genes.";  
 RL Mamm. Genome 6:103-110(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kuge S., Ihara S., Watanabe E., Watanabe M., Takishima K., Suga T.,  
 RA Mamaiya G., Kawakami M.;  
 RA Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-  
 CC DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,  
 CC BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF  
 CC THE ANTIBODY.  
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.  
 CC -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST  
 CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI  
 CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND  
 CC LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNALE  
 CC SPACE OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE  
 CC PROTEIN.  
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; S42294; AAB19343.1; -.  
 CC EMBL; U09016; AAA82010.1; -.  
 CC EMBL; U09013; AAA82010.1; JOINED.  
 CC EMBL; U09014; AAA82010.1; JOINED.  
 CC EMBL; U09015; AAA82010.1; JOINED.  
 CC EMBL; D11440; BAA02005.1; -.  
 CC HSP; P08661; IPDO.  
 CC MGD; MGI:96924; Mbl2.  
 CC InterPro; IPR000087; Collagen.  
 CC InterPro; IPR001304; lectin\_c.  
 CC Pfam; PF00059; lectin\_c; 1.  
 CC SMART; SM00034; CLECT; 1.  
 CC PROSITE; PS00615; C-TYPE LECTIN\_1; 1.  
 CC PROSITE; PS50041; C-TYPE LECTIN\_2; 1.  
 CC Lectin; Hydroxylation; Liver; Glycoprotein; Mannose-binding; Membrane;  
 CC Calcium; Collagen; Repeat; Signal.  
 CC SIGNAL 1 18  
 FT CHAIN 19 244 MANNOSE-BINDING PROTEIN C.  
 FT DOMAIN 38 96 COLLAGEN-LIKE.  
 FT DOMAIN 149 242 C-TYPE LECTIN (SHORT FORM).  
 FT MOD\_RES 43 43 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 58 58 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 69 69 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 78 78 HYDROXYLATION (POTENTIAL).  
 FT MOD\_RES 81 81 HYDROXYLATION (POTENTIAL).  
 FT DISULFID 29 29 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 34 34 INTERCHAIN (BY SIMILARITY).

FT DISULFID 34 34 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 151 240 BY SIMILARITY.  
FT DISULFID 218 232 BY SIMILARITY.  
FT CONFLICT 3 3 I -> L (IN REF. 1).  
FT CONFLICT 15 15 V -> A (IN REF. 1).  
SQ SEQUENCE 244 AA; 25957 MW; 49AE84E2290DEB0A CRC64;

Query Match 53.5%; Score 83; DB 1; Length 244;  
Best Local Similarity 59.3%; Pred No. 4.4e-05;  
Matches 16; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 EKVEMYTDGKNDRNCLQSLAICEF 27  
| | | : | | | | | | | | | |  
Db 216 EDCVVILGNGKNDVPDCSDSFLAICEF 242

Search completed: July 3, 2002, 12:40:43  
Job time: 351 sec

Result No.	Query	Score	Match	Length	DB	ID	Description
✓ 1	✓ 1	352.5	23.8	369	2	S33603	surfactant protein
✓ 2	✓ 2	345	23.2	301	2	A53570	collectin-43 - bov
3		337	22.7	375	1	A45225	pulmonary surfacta
4		333	22.4	374	1	A42046	surfactant protein
5		328.5	22.1	371	1	JN0450	conglutinin precu
6		328.5	22.1	371	2	I45878	conglutinin - bov
7		305.5	20.5	247	1	LNRBPS	pulmonary surfacta
8		290	19.5	238	1	LNRMSA	mannose-binding le
9		290	19.5	244	1	LNRMSM	mannose-binding le
10		289.5	19.5	248	1	LNRHMC	mannose-binding le
11		287	19.3	248	1	LNRHUP1	pulmonary surfacta
12		285.5	19.2	248	1	LNRHPS	pulmonary surfacta
13		284.5	19.2	248	2	I51921	pulmonary surfacta
14		284	19.1	248	1	LNRHUP6	pulmonary surfacta
15		280	18.9	248	1	LNRDPS	pulmonary surfacta
16		279	18.8	244	1	LNRTHC	mannose-binding le
17		277	18.7	248	1	LNRTPS	pulmonary surfacta
18		273	18.4	742	2	JC7595	scavenger receptor
19		270.5	18.2	239	1	LNRMSA	mannose-binding le
20		269.5	18.2	248	2	A48853	pulmonary surfacta
21		207	13.9	618	2	S32436	collagen alpha 2(I
22		202	13.8	645	2	D90782	probable tail fibe
23		205	13.8	645	2	H85642	probable tail fibe
24		203	13.7	688	2	A53330	collagen alpha 2(I
25		196	13.2	775	2	A53893	collagen alpha 1(I
26		193	13.0	1549	2	I48103	type VII collagen
27		192.5	13.0	2944	2	A54849	collagen alpha 1(V
28		191.5	12.9	1366	1	CGHU25	collagen alpha 2(I
29		191	12.9	677	2	S32395	collagen alpha 2(I

Query Match	22.7%;	Score 337;	DB 1;	Length 375;	
Best Local Similarity	35.2%;	Pred. No. 2.4e-20;			
Matches	86;	Conservative 28;	Mismatches 110;	Indels 20;	Gaps 6;
46	PGPKGDDGEKGDPEEGEKHGKVGWPKIGKELGDMGRDNITGKTPGK-----	96			
dbb	:    :				
135	PGPKGEAPGKEVGAPCMQGSAGARGLAGPKGERSVPGTGAAGSAGAMPQGS	194			
QY	:    :				
97	---KGDGKEKGLLGPGEKGKAGI--VCDGGRYKRFVQGLDISIARLKTSMKFKVNV--	149			
dbb	:    :				
195	PGARGPKLKGDKGIPGDKGAKGESGLPDVASLRQGVFALQGVQHLQAAAFSQYKKYELF	254			
QY	:    :				
150	AGRIETEKEKFFYIVQBEKNYRESLTHCRIRGGMLAMPDEAANTLIAD-VYAKSGFFRVF	208			
dbb	:    :				
255	PNQGSVGEKIFKTAGVKEFTEAQLLCTQAGGOLASPRSAENNAQQILVYAKNE--AAF	312			
QY	:    :				
209	IGVNDLERGEQYMTDNTPLQYNNWNEGSDPYCHDCDCEYMLSSGRVNDTECHLTMFY	268			
dbb	:       :       :				
313	LSMTDSKTEGKFTVPTGESLV-YSNWAPGPNDDGSDCEVFETNGKNDRACGERLV	371			
QY	:       :       :				
269	VCEF 272				
dbb					
372	VCEF 375				
db					



A>Title: The cDNA cloning of conglutinin and identification of liver as a primary site  
A:Reference number: S33235; MUID:93277452  
A:Accession: S33235  
A:Molecule type: mRNA  
A:Residues: 1-172, 'H','L',174-217,'A','219-271','V',273-371 <LUJ>  
A:Cross-references: EMBL:X71774; NID:G395267; PIDN:CAAS0665.1; PID:G395268  
A:Experimental source: liver  
R:Lee, Y.M.; Leiby, K.R.; Allar, J.; Paris, K.; Lerch, B.; Okarma, T.B.  
J. Biol. Chem. 266, 2715-2723, 1991  
A>Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin family  
A:Reference number: A23740; MUID:91131556  
A:Accession: A23740  
A:Molecule type: protein  
A:Residues: 21-209,'S','211-371 <LEE>  
R:Kawasaki, N.; Yokota, Y.; Kawasaki, T.  
Arch. Biochem. Biophys. 305, 533-540, 1993  
A>Title: Differentiation of conglutination activity and sugar-binding activity of cor  
A:Reference number: S36879; MUID:93384312  
A:Accession: S36879  
A:Molecule type: protein  
A:Residues: 21-54;'75'-86,'X','88-89,'X','91,'X',93-94;208-209,'X',211-227 <KAW>  
A:Experimental source: serum  
R:Lü, J.; Wiedemann, H.; Holmskov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.  
Eur. J. Biochem. 215, 793-799, 1993  
A>Title: Structural similarity between lung surfactant protein D and conglutinin. Two  
A:Reference number: S35044; MUID:93359905  
A:Accession: S35044  
A:Molecule type: protein  
A:Residues: 75-86,'X','88-89,'X','91,'I' <LOA>  
A:Experimental source: lung  
R:Young, N.M.; Leon, M.A.  
Biochem. Biophys. Res. Commun. 143, 645-651, 1987  
A>Title: The carbohydrate specificity of conglutinin and its homology to proteins in  
A:Reference number: A29416; MUID:87184551  
A:Contents: annotation  
A:Malhotra, R.; Laursen, S.B.; Willis, A.C.; Sim, R.B.  
Biochem. J. 293, 15-19, 1993  
A>Title: Research Communication. Localization of the receptor-binding site in the col  
A:Reference number: S34054; MUID:93319501  
A:Contents: annotation  
R:Liou, L.S.; Saxty, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sas  
J. Immunol. 153, 173-180, 1994  
A>Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship  
A:Reference number: I46010; MUID:94267222  
A:Accession: I46010  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-371 <LIO>  
A:CROSS-references: EMBL:U06860; NID:Q507183; PIDN:AAB60624.1; PID:G514256  
C:Comment: This protein mediates the agglutination of erythrocytes with antibody and  
C:Comment: This protein is a Ca<sup>2+</sup>-dependent serum lectin specific for N-acetylglucosa  
C:Genetics:  
A:Gene: CGNI  
A:Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1  
C-Superfamily: pulmonary surfactant protein D; C-type lectin homology  
C:Keywords: calcium binding; glycoprotein; hydroxylysine; hydroxyproline  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-371/Product: conglutinin #status predicted <MAT>  
F:46-214/Region: collagen-like  
F:F75-371/Product: conglutinin-N #status predicted <MA>  
F:248-369/Domain: c-type lectin homology <ICH>  
F:E63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #statu  
F:E63,87,99,135,141,159,162,198,210/Modified site: 5-hydroxylsine (Lys) #status exper  
F:E78,96,108,111,129,132,147,153,171,195/Modified site: 4-hydroxyproline (Pro) #status  
F:337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 22.1%; Score 328.5; DB 1; Length 371;  
Best Local Similarity 29.5%; Pred. No. 1.2e-19;  
Matches 85; Conservative 35; Mismatches 95; Indels 73; Gaps 10;

OY 46 PPKGDGDEKDP-----GEGRKHGVGMGPKGIGELGDMGRGNITGK-----TGP 93  
||||| : | : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :

Db 96 PGKGTGTRPGMPGAGREGPSGKQSGMPPFTGPKGTGPKGVGAGPQGGP 155  
 QY 94 IGKKGDKGKGLLTPGE-----KKGAG- 116  
 Db 156 SGLKGEKAGPGETGAPGRAGVTGPGSAGTQPGSARGPGLKGRDGPGETGAKGESGL 215  
 QY 117 -----TDCGGRYKRVGQLDSTARKTSMKFKVNVJAGIRETEKFFYIVQEE 166  
 Db 216 AEVNALKQRTILD-GHLRRFQ-----AFSQYKKAFLFPDQAVG-----EKIFKTAGAV 265  
 QY 167 KNYEESLTHCRIRGGLAMPKDEAANTLIADYVAKSGFFRVFGVNDLEREGQYMTDNT 226  
 Db 266 KYSYDAEQICREAKQJASPRSAEAEVQMV-RAQEKAYLSMNDISTEGRETYPTGE 324  
 QY 227 PLQYSNWNEGEP--SDPYGHEDCVEMLSGRWNTDTECHLTMTFVCEEF 272  
 Db 325 ILV-YSNWDGEPNNSDEGQENCVELFPDGKNDVPCSKQLLVICEF 371  
 RESULT 6  
 I45878  
 conglutinin - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 19-Dec-1997 #sequence\_revision 19-Dec-1997 #text\_change 20-Aug-1999  
 C:Accession: I45878  
 R:Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry  
 Gene 141, 277-281, 1994  
 A>Title: Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization of  
 A:Reference number: I45878; MUID:94215917  
 A:Accession: I45878  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-371 <LCH>  
 A:Cross-references: GB:L18871; NID:G495012; PIDN:AAA20126.1; PID:G495013  
 C:Superfamily: pulmonary surfactant protein D; C-type lectin homology  
 F:248-369/Domain: C-type lectin homology <LCH>

Query Match 22.1%; Score 328.5; DB 2; Length 371;  
 Best Local Similarity 29.5%; Pred. No. 1.2e-19;  
 Matches 85; Conservative 35; Mismatches 95; Indels 73; Gaps 10;  
 QY 46 PGKGGDDGKGGP-----GEECKHKGKGMGPKGKIGKELGDMGRGNIGK-----TGP 93  
 Db 96 PGKGTGTRPGMPGAGREGPSGKQSGMPPFTGPKGTGPKGVGAGPQGGP 155  
 QY 94 IGKKGDKGKGLLTPGE-----KKGAG- 116  
 Db 156 SGLKGEKAGPGETGAPGRAGVTGPGSAGTQPGSARGPGLKGRDGPGETGAKGESGL 215  
 QY 117 -----TDCGGRYKRVGQLDSTARKTSMKFKVNVJAGIRETEKFFYIVQEE 166  
 Db 216 AEVNALKQRTILD-GHLRRFQ-----AFSQYKKAFLFPDQAVG-----EKIFKTAGAV 265  
 QY 167 KNYEESLTHCRIRGGLAMPKDEAANTLIADYVAKSGFFRVFGVNDLEREGQYMTDNT 226  
 Db 266 KYSYDAEQICREAKQJASPRSAEAEVQMV-RAQEKAYLSMNDISTEGRETYPTGE 324  
 QY 227 PLQYSNWNEGEP--SDPYGHEDCVEMLSGRWNTDTECHLTMTFVCEEF 272  
 Db 325 ILV-YSNWDGEPNNSDEGQENCVELFPDGKNDVPCSKQLLVICEF 371

RESULT 7  
 LNRPBS  
 pulmonary surfactant protein A precursor - rabbit  
 N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999  
 C:Accession: A29931  
 R:Boygaram, V.; Qing, K.; Mendelson, C.R.  
 J. Biol. Chem. 263, 2939-2947, 1988  
 A>Title: The major apoprotein of rabbit pulmonary surfactant. Elucidation of primary seq

A:Reference number: A29931; MUID:88139348  
 A:Accession: A29931  
 A:Molecule type: mRNA  
 A:Residues: 1-247 <BCG>  
 A:Cross-references: GB:J03542; NID:G165705; PIDN:AAA31465.1; PID:G165706  
 A:Note: 12-Pro was also found  
 A:Note: two species of mRNA, which appear to be transcribed from a single gene, could  
 A:Note: the amino end of the mature protein is blocked  
 C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lower  
 C:Comment: This protein is a sialoglycoprotein synthesized by alveolar type II cells.  
 C:Comment: dependent on the presence of calcium ions.  
 C:Superfamily: mannose-binding lectin; C-type lectin homology  
 C:Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; gl  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-247/Product: pulmonary surfactant protein A #status predicted <MAT>  
 F:27-99/Region: collagen-like  
 F:126-245/Domain: C-type lectin homology <LCH>  
 F:16/Modified site: acetylated amino end (Ser) (in mature form) #status predicted  
 F:206/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 20.6%; Score 305.5; DB 1; Length 247;  
 Best Local Similarity 31.2%; Pred. No. 5.9e-18;  
 Matches 85; Conservative 30; Mismatches 104; Indels 53; Gaps 9;  
 QY 25 SLGLDIDSRPTAEVCATHI---SPGPKGDDGKPGEGKHKGVGRMCPKGIKGLGD 81  
 Db 5 SLALTLSAPASDCTDVCVIGSGPGIPGPGSHGLPGRDGVKGDGPGPGMPPGG 64  
 QY 82 M-----GDRGNIGKTPGPKGKGEKGLGIPGEGKAGTVCDCGRYKRVGQLD----- 132  
 Db 65 MPGLPGRDGLIGAPGVPGERGDKGE-----PGERGPPG-----LPAYLDEELQA 108  
 QY 133 -----ISIAKTKSMKFKVNVJAGIRETEKFFYIVQEEKNYRESLTHCRIRG 180  
 Db 109 TLHELRLHQAQSLVLSLOGSMKAV-----GEKIFSTGQSVNDFATREVCARAG 158  
 QY 181 GLMLPDKAANTLIADYVAKSGFFRVFGVNDLEREGQYMTDNTPLQYSNWNEGEP 240  
 Db 159 GRIAPVPSLENEAIAIVKERNYAYLAGBPTAGDFYLLGDGPVNYINWYPGEPR 216  
 QY 241 DPYGHEDCVEMLSGRWNTDTECHLTMTFVCEEF 272  
 Db 217 G-QGREKCVEMTGDKNWDCNCLQYRLVICEF 247

RESULT 8  
 LNRTMA  
 mannose-binding lectin A precursor - rat  
 N:Alternate names: serum mannan-binding protein  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 22-Jun-1999  
 C:Accession: B24791; A29556; A27799  
 R:Drickamer, K.; Dordal, M.S.; Reynolds, L.  
 J. Biol. Chem. 261, 6878-6887, 1986  
 A>Title: Mannose-binding proteins isolated from rat liver contain carbohydrate-recogn  
 A:Reference number: A24791; MUID:86196130  
 A:Accession: B24791  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <DR1>  
 R:Drickamer, K.; McCreary, V.  
 J. Biol. Chem. 262, 2582-2589, 1987  
 A>Title: Exon structure of a mannose-binding protein gene reflects its evolutionary r  
 A:Reference number: A29556; MUID:87137502  
 A:Accession: A29556  
 A:Molecule type: DNA  
 A:Residues: 1-155, 'K', 157-238 <DR2>  
 A:Cross-references: GB:M14104; GB:M14105; NID:G205259; PIDN:AAA98781.1; PID:G205261  
 A:Note: the codon AAG for residue 156 is inconsistent with the authors' statement tha  
 R:Ikeda, K.; Sannoh, T.; Kawasaki, N.; Kawasaki, I.  
 J. Biol. Chem. 262, 7451-7454, 1987  
 A>Title: Serum lectin with known structure activates complement through the classical

A:Title: cDNAs and deduced amino acid sequences of subunits in the binding component  
A:Reference number: A27799; MUID:87222358  
A:Accession: A27799  
A:Molecule type: protein  
A:Residues: 18-42 <IKK>  
C:Comment: Mannose-binding lectins are opsonins that are important in host defense against bacteria and fungi. This plasma protein binds mannose and N-acetylglucosamine and can activate complement. The molecule consists of approximately 20 identical chains linked by disulfide bonds.  
C:Genetics:  
A:Introns: 51/1; 90/1; 115/1  
C:Superfamily: mannose-binding lectin; C-type lectin homology  
C:Keywords: acute phase; calcium; hydroxylysine; hydroxyproline; lectin; liver; plasma  
F:1-17/Domain: signal sequence #status predicted <SIG>  
F:18-238/Product: mannose-binding lectin A #status experimental <MAT>  
F:36-88/Region: collagen-like  
F:85-87/Region: cell attachment (R-G-D) motif  
F:118-234/Domain: C-type lectin homology <LCH>  
F:61.67.73/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:79.82/Modified site: lysine derivative (Lys) (probably 5-hydroxylysine) #status experimental

Query Match 19.5%; Score 290; DB 1; Length 238;  
Best Local Similarity 28.7%; Pred. No. 1.1e-16;  
Matches 75; Conservative 49; Mismatches 105; Indels 32; Gaps 8;

QY 14 ILVLFLQLQSLGLDIDSPRAEVCATHHTISPGKGDGEGKDPGEGKHKVGRMGPK 73  
DQ 14 ILVLFLQLQSLGLDIDSPRAEVCATHHTISPGKGDGEGKDPGEGKHKVGRMGPK 73  
Db 6 LVLVLLCVSVSSSQ--TCETILKTS--VIACGRDGRGPKGEGKGPQ-----GLR 55

QY 74 GIKGELGMDGRNGTIGTIGKKGDKGKGLLIGPCKGKAGTCDGGRYKRVFGOLDI 133  
DQ 74 GIKGELGMDGRNGTIGTIGKKGDKGKGLLIGPCKGKAGTCDGGRYKRVFGOLDI 133  
Db 56 GLOGPGKLGPGSVGAPGSGQPKQKQDGR-----DSRAIEVKLANMEA 100

QY 134 SIARLTSKFKVNIAGI--RETEKYFYIYQVEKYNRESLTHCRIRGGLAMPKDEAA 191  
DQ 134 SIARLTSKFKVNIAGI--RETEKYFYIYQVEKYNRESLTHCRIRGGLAMPKDEAA 191  
Db 101 EINTLAKSLDELTLNLFHAFSGKSGKGFVTNHERMPFSGKALCSLSELGTVAIIPNAEE 160

QY 192 NTLIADYVAKSGFRFVIGVNDLEREGQYMTDTNTPLOYSNNWNGEPPSDYHECDVEM 251  
DQ 192 NTLIADYVAKSGFRFVIGVNDLEREGQYMTDTNTPLOYSNNWNGEPPSDYHECDVEM 251  
Db 161 NKAIQF-VAKYS---AFGLGIDTEVGEQFMVVTGGRL-TYSNWKDEPNDRGSDCVTI 215

QY 252 LSSGRNNDTECHLTMYFVCEP 272  
DQ 252 LSSGRNNDTECHLTMYFVCEP 272  
Db 216 VDNLGLNDISQASHATVACEF 236

RESULT 9  
LNHSMC  
mannose-binding lectin C precursor - mouse  
N:Alternate names: R-reactive factor P28a  
C:Species: Mus musculus (house mouse)  
C:Date: 18-Jun-1993 #sequence\_revision 20-Feb-1998 #text\_change 16-Jun-2000  
R:Accession: I48651; B46466; A42574; C42574  
R:Sastry, R.; Wang, J.S.; Brown, D.C.; Ezekowitz, R.A.; Tauber, A.I.; Sastry, K.N.  
Mamm. Genome 6, 103-110, 1995  
A:Title: Characterization of murine mannose-binding protein genes Mb11 and Mb12 reveals  
A:Reference number: I48650; MUID:95284466  
A:Accession: I48651  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-244 <SAS1>  
A:Cross-references: EMBL:U090116; NID:q773286; PIDN:AA82010.1; PID:q773288  
R:Sastry, K.; Zahedi, K.; Lelias, J.M.; Whitehead, A.S.; Ezekowitz, R.A.  
J. Immunol. 147, 692-697, 1991  
A:Title: Molecular characterization of the mouse mannose-binding proteins. The mannose-b  
A:Reference number: A46466; MUID:91302823  
A:Accession: B46466  
A:Molecule type: mRNA  
A:Residues: 1-2, '4, '14, 'A, '16-244 <SAS2>  
A:Cross-references: GB:542294; NID:q233017; PIDN:AAB19343.1; PID:q233018  
A:Experimental source: Inbred CBA/J, acute phase liver library, ptz 19 vector  
A:Note: Sequence extracted from NCBI backbone (NCBIN:42294, NCBIP:42295)  
R:Kuge, S.; Ihara, S.; Watanabe, E.; Watanabe, M.; Takishima, K.; Suga, T.; Mamiya, G.;  
Biochemistry 31, 6943-6950, 1992

A:Title: cDNAs and deduced amino acid sequences of subunits in the binding component  
A:Reference number: A42574; MUID:92345256  
A:Accession: A42574  
A:Molecule type: mRNA; protein  
A:Residues: 1-244 <KUG>  
A:Cross-references: GB:D11440; NID:q220585; PIDN:BA02005.1; PID:q220586  
A:Experimental source: BALB/c, liver  
A:Note: Sequence extracted from NCBI backbone (NCBIP:110137)  
A:Note: parts of the sequence, including the amino end of the mature protein, were co  
A:Accession: C42574  
A:Molecule type: protein  
A:Residues: 19, 'X, '21-28, 'X', '30-32; 72-77, 'H', '79-80; 'G', '177-185; 187-189, 'H', '191-198 <K  
A:Note: Source is serum of ICR mice; differences may be allotypic  
C:Genetics:  
A:Gene: Mb12  
A:Introns: 59/1; 98/1; 121/1  
C:Superfamily: mannose-binding lectin; C-type lectin homology  
C:Keywords: calcium; endoplasmic reticulum; Golgi apparatus; homohexameric; hydroxyprol  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-244/Product: mannose-binding lectin C #status experimental <MAT>  
F:38-94/Region: collagen-like  
F:124-240/Domain: C-type lectin homology <LCH>  
F:29,34/disulfide bonds: interchain #status predicted  
F:69/Modified site: 4-hydroxyproline (Pro) #status experimental

Query Match 19.5%; Score 290; DB 1; Length 244;  
Best Local Similarity 28.8%; Pred. No. 1.1e-16;  
Matches 80; Conservative 56; Mismatches 100; Indels 42; Gaps 11;

QY 1 MNGFASILRRNQFILLVFLQLQSLGLDIDSPRAEVCATHHTIS--PGPKGDGEGKDP 58  
DQ 1 MNGFASILRRNQFILLVFLQLQSLGLDIDSPRAEVCATHHTIS--PGPKGDGEGKDP 58  
Db 1 MSITFSL-----LLCVTVVYAEITLLEGVQNSCPVWTCSSPGLNGEPGKGRDGAKEK 55

QY 59 GEEGKHGKVRMGPKGKIGELGMDGRNGTIGTIGKKGDKGKGLLIGPCKGKAGTCDGGRYKRVFGOLDI 118  
DQ 59 GEEGKHGKVRMGPKGKIGELGMDGRNGTIGTIGKKGDKGKGLLIGPCKGKAGTCDGGRYKRVFGOLDI 118  
Db 56 GEPGQ-----GLRGLQPPGKVGTPGPGPKGLGAVGPKGDRGDA----- 97

QY 119 CDGRYKRF-VGQDLSITARLTSKFKVNIAGI--RETEKYFYIYQVEKYNRESLTH 175  
DQ 119 CDGRYKRF-VGQDLSITARLTSKFKVNIAGI--RETEKYFYIYQVEKYNRESLTH 175  
Db 98 -----EFDTSEIDSEIAALSELRAIRNWLIFSLSEKVKCKFYVSSVKMSLDRVKAL 150

QY 176 CRIRGGLAMPKDEANTLIADYVAKSGFRFVIGVNDLEREGQYMTDTNTPLOQ-NYSNW 234  
DQ 176 CRIRGGLAMPKDEANTLIADYVAKSGFRFVIGVNDLEREGQYMTDTNTPLOQ-NYSNW 234  
Db 151 CSEFGSGVATPRNAENSAI-QKVAKD---IAYLGITDVRVSGS--FEDLTGNRVRYTNW 204

QY 235 NGEPSDPYGHEDCYEMLSRGRWNTDECHLTMYFVCEP 272  
DQ 235 NGEPSDPYGHEDCYEMLSRGRWNTDECHLTMYFVCEP 272  
Db 205 NDGPENNTGDEDCVVILGNGKWDVPCSDSFLAICEF 242

RESULT 10  
LNHSMC  
mannose-binding lectin precursor [validated] - human  
N:Alternate names: mannan-binding protein  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1989 #sequence\_revision 30-Sep-1991 #text\_change 08-Dec-2000  
C:Accession: J10115; S05641; A34978; J10027; JX0319; PC2188; A32266  
R:Sastry, K.; Herman, G.A.; Day, L.; Deignan, E.; Bruns, G.; Morton, C.C.; Ezekowitz,  
J. Exp. Med. 170, 1175-1189, 1989  
A:Title: The human mannose-binding protein gene. Exon structure reveals its evolution  
A:Reference number: J10115; MUID:90010778  
A:Accession: J10115  
A:Molecule type: DNA  
A:Residues: 1-248 <SAS>  
A:Cross-references: EMBL:X15422; NID:q34486; PIDN:CAA33462.1; PID:q34487  
R:Taylor, M.E.; Brickell, P.M.; Craig, R.K.; Summerfield, J.A.  
Biochem. J. 262, 763-771, 1989  
A:Title: Structure and evolutionary origin of the gene encoding a human serum mannose  
A:Reference number: S05641; MUID:90073571  
A:Accession: S05641  
A:Molecule type: DNA  
A:Residues: 1-248 <TAY>

A:Cross-references: EMBL:X15954; NID:g34480; PIDN:CAA34079.1; PID:g1212951

A:Accession: A34978

A:Molecule type: protein

A:Residues: X',22-24,'X',26,'X',28-31,'X',33-34,'X',36,'XXXX',41-50 <TAY2>

R:Ekowitz, R.A.B.; Day, L.E.; Herman, G.A.

J. Exp. Med. 167, 1034-1046, 1988

A:Title: A human mannose-binding protein is an acute-phase reactant that shares sequence

A:Reference number: J10027; MUID:88171281

A:Accession: J10027

A:Molecule type: mRNA

A:Residues: 1-2,'C',4,'T',8,'S',10-57,'R',59-60,'G',63-106,'PQCLRK',113,'SSANRNGTYQ',1

R:Kurata, H.; Sannoh, T.; Kozutsumi, Y.; Yokota, Y.; Kawasaki, T.

J. Biochem. 115, 1148-1154, 1994

A:Title: Structure and function of mannose-binding proteins isolated from human liver and

A:Reference number: JX0319; MUID:95073978

A:Accession: JX0319

A:Molecule type: protein

A:Residues: 1-248 <KUR1>

A:Accession: PC2188

A:Molecule type: mRNA

A:Residues: 1-20 <KUR2>

A:Experimental source: liver and serum

C:Comment: Mannose-binding lectins are opsonins that are important in host defense against

C:Comment: This protein is a Ca<sup>2+</sup>-requiring animal lectin specific for mannose and N-acetyl

C:Genetics:

A:Gene: GDB:MBL

A:Cross-references: GDB:120167; OMIM:154545

A:Map position: 10q11.2-10q11.2

A:Introns: 63/1; 102/1; 125/1

C:Superfamily: mannose-binding lectin; C-type lectin homology

C:Keywords: acute phase; calcium binding; endoplasmic reticulum; Golgi apparatus; hydrox

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-248/Product: mannose-binding lectin #status experimental <MAT>

F:42-99/Region: collagen-like

F:128-244/Domain: C-type lectin homology <LCH>

F:47,73,79,82,88/Modified site: 4-hydroxyproline (Pro) (partial) #status experimental

Query Match 19.5%; Score 289.5; DB 1; Length 248;

Best Local Similarity 31.1%; Pred. No. 1.2e-16;

Matches 75; Conservative 37; Mismatches 100; Indels 29; Gaps 7;

QY 34 PTAECVATHIS--PGPKGDGEGKGPGEKGKGVGRMGPKIGKIGELGDMGRGNIGKT 91

Db 33 PAVIACSSPFGINGFGKGRDGTGKEGEPQ-----GLRGLOGPPGKLGPPGPNGPS 85

QY 92 GPTEGKKGKGGKGLLGPGEKAGTVCDGGRYKFGVQGLDISIARLTKSMKFKVNIAG 151

Db 86 GSPGPKGQKDPG-----KSPDGDSSLAASERK---ALQTEMARIKKWLITSLG--- 131

QY 152 IRETEKFFYVQEEKYRESLTHCRIRGGMAMPKDEAANTLIADYVAKSGFFRYFGV 211

Db 132 -KQVGNKFLNGELMTFEKVALCVKQASVATPRNAENGAIONLIKE-----EAFGI 186

QY 212 NDLERGGYMTDNTPLQNYSNWNEGSPDYGHEDCVEMLSGGWNDTECHLTMYFVCE 271

Db 187 TDEKTEGQFVLTGNRL-ITYWNNGEPNAGSDEDCVLLKNGWQNDVPCSTSHLAVCE 245

QY 272 F 272

Db 246 F 246

RESULT 11

LNHUP1

Pulmonary surfactant protein A precursor (clone 1A) - human

N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999

A:Accession: B25720

R:Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.; Recny, M.; Sultzman, L.

J. Biol. Chem. 261, 9029-9033, 1986

A:Title: Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfacta

A:Reference number: A25720; MUID:86250832

A:Accession: B25720

A:Molecule type: mRNA

A:Residues: 1-248 <FLO>

A:Cross-references: GB:K03475

A:Note: part of the sequence was confirmed by protein sequencing

A:Note: the amino end of the mature protein, which was not identified, is partially a

A:Note: clones corresponding to two different proteins were sequenced. Cotranslationa

C:Genetics:

A:Gene: GDB:SFTPA1; SFTPL; SP-A; SP-A1

A:Cross-references: GDB:119593; OMIM:178630

A:Map position: 10q22-10q23

C:Superfamily: mannose-binding lectin; C-type lectin homology

C:Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; gl

F:1-248/Domain: signal sequence #status predicted <SIG>

F:127-246/Product: pulmonary surfactant protein A #status predicted <MAT>

F:21/Modified site: acetylated amino end (Glu) (in mature form) #status predicted

F:30,33,36,42,54,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status p

F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 19.3%;

Best Local Similarity 28.2%; DB 1; Length 248;

Matches 75; Conservative 44; Mismatches 117; Indels 30; Gaps 9;

QY 14 ILLVFLLIQISLGLDIDSRPTAEVCATHTISPGPKGDGEGKGPGEKGKGVGRMGPK 73

Db 6 LALTLLMAASGAACE-----KDCVCG---SPGIPGTPSGHGLPGDRGLKDGPP 57

QY 74 GIKELGDM-----GGRNIGTGPIGKDKGKGLLGPGEKAGTVCDGGRYKRVG 129

Db 58 GPMGPGETPCPPGNNGLPAGPVGPERGEKGE-----PGERGPGPLPAHLDE----- 105

QY 130 OLD:SIARLTKSMKFKVNIAG---GIRETEKFFYVQEEKYRESLTHCRIRGGMAMP 186

Db 106 ELQATLHDFHQLIOTRGLSLOGSIMTVGKVFSSNGQSITDAIOEACARAGRIATVP 165

QY 187 KDEAANTLIADYVAKSGFFRYFGVNDLREGGYMTDNTPLQNYSNWNEGSPDYGHE 246

Db 166 NPENEATIASFKVKNYI-AYVGLTEGSPGDFRYSDGTPV-NYTNWYRGEPAG-RGKE 222

QY 247 DCVEMLSGGWNDTECHLTMYFVCEP 272

Db 223 QCVEMYTDGQWDRNCLYSRLTICEF 248

RESULT 12

LNHUPS

Pulmonary surfactant protein A precursor (genomic clone) - human

N:Alternate names: alveolar proteinosis protein; pulmonary surfactant 32K apoprotein;

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999

A:Accession: A24622; A43628

R:White, R.T.; Damm, D.; Miller, J.; Spratt, K.; Schilling, J.; Hawgood, S.; Benson,

Nature 317, 361-363, 1985

A:Title: Isolation and characterization of the human pulmonary surfactant apoprotein

A:Reference number: A24622; MUID:86014366

A:Accession: A24622

A:Molecule type: DNA

A:Residues: 1-248 <WHI>

A:Cross-references: GB:M30838; NID:g190564; PIDN:AAA36510.1; PID:g190565

A:Note: the sequence in GenBank entry HUMPSAP, release 109.0, (PID:g190565) has the c

A:Note: four nucleotide differences, producing amino acid differences at positions 45

R:Hagman, H.P.; White, R.T.; Schilling, J.; Lau, K.; Benson, B.J.; Golden, J.; Hawg

Am. J. Physiol. 257, 1421-1429, 1989

A:Title: Studies of the structure of lung surfactant protein SP-A.

A:Reference number: A43628; MUID:90119861

A:Accession: A43628

A:Molecule type: protein

A:Residues: 143-150;220-240;243-248 <HAA>

C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lower

C:Comment: This protein is a sialoglycoprotein synthesized by alveolar type II cells.

pendent on the presence of calcium ions.

C:Genetics:  
A:Gene: GDB:SFTPAL; SFTPL; SP-A; SP-AL  
A:Cross-references: GDB:119593; OMIM:178630  
A:Map position: 10q22-10q23  
A:Introns: 58/1; 98/1; 124/1  
C:Superfamily: mannose-binding lectin; C-type lectin homology  
C:Keywords: alveolar proteinosis; calcium; gaseous exchange; glycoprotein; hydroxylysine  
F:1-20/Domain: signal sequence #status predicted <STG>  
F:21-248/Product: pulmonary surfactant protein A #status predicted <MAT>  
F:28-100/Domain: collagenous #status predicted <COL>  
F:127-246/Domain: C-type lectin homology <LCH>  
F:26/Disulfide bonds: interchain #status experimental  
F:30,33,36,42,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status predicted  
F:51,86/Modified site: 5-hydroxylysine (Lys) #status predicted  
F:155-246,224-238/Disulfide bonds: #status experimental  
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 19.2%; Score 285.5; DB 1; Length 248;  
Best Local Similarity 30.6%; Pred. No. 2.7e-16;  
Matches 74; Conservative 37; Mismatches 112; Indels 19; Gaps 8;  
QY 38 VCATHI---SPGKGDGEGKGVKRMGPKIGKIGELGDMG-DRGNIGTKTP 93  
DB 19 VEEVDVGVSGPIPTPSHGLPGRGRGLKGLDGGPPGPPGEMPCPPGNDGLPGA 78  
QY 94 ICKGDKGKGLLIGFGEKAGTVCDCGRYKRFVGLDLSIARLTKSMKFKVKNVIA--- 150  
DB 79 PGIPGECGEK---GEPGPGPLPAHLDE-----ELQATLHDFRQILQTRGALSLOG 129  
QY 151 GIRETEEFYIVQEEKYRESLHCRIRGMLAMPKDAANTLIADYVAKSGFFRVTIG 210  
DB 130 SMTVGEKVFSSNGQSITDAIQACARAGGRIAVRNPENEAIASFVKKYNTY-AVVG 188  
QY 211 VNDLRECOYMTDNTPLQNYNNWNEGEPDYPGHEDCVEMLSGRWNTDECHLTMFYWC 270  
DB 189 LTEGSPGDFRYSDGTPV-NYTNWYRGEPAK-RGKQCQVEMTQGNDRNCLYSRLTIC 246  
QY 271 EF 272  
DB 247 EF 248

# RESULT 13

151921  
Pulmonary surfactant-associated protein A1 - human  
N:Alternate names: SP-A1  
C:Species: Homo sapiens (man)  
C:Accession: I51921  
R:Katyal, S.L.; Singh, G.; Locker, J.  
Am. J. Respir. Cell Mol. Biol. 6, 446-452, 1992  
A:Title: Characterization of a second human pulmonary surfactant-associated protein SP-A  
A:Reference number: I51921; MUID:92198680  
A:Accession: I51921  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-248 <RES>  
A:Cross-references: GB:M68519; NID:g338048; PIDN:AAA60319.1; PID:g338049  
C:Genetics:  
A:Gene: GDB:SFTPAL; SFTPL; SP-A; SP-AL  
A:Cross-references: GDB:119593; OMIM:178630  
A:Map position: 10q22-10q23  
A:Introns: 58/1; 98/1; 124/1  
C:Superfamily: mannose-binding lectin; C-type lectin homology  
F:127-246/Domain: C-type lectin homology <LCH>

Query Match 19.2%; Score 284.5; DB 2; Length 248;  
Best Local Similarity 31.2%; Pred. No. 3.2e-16;  
Matches 70; Conservative 37; Mismatches 110; Indels 7; Gaps 5;

QY 53 GEKGDPEGEKHKYGRMGPKIGKIGELGDMGRGNIGTKT-PGKKGDKGKGLLIGPE 111  
DB 28 GSPGIPOTPGSHGLPGRDRGVKGDGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 87  
QY 112 KGKAGTVCDCGRYKRFVGLDLSIARLTKSMKFKVKNVIA---GIRETEEFYIVQEEK 168  
DB 86 KGEAGERGPPGLPAHLDEBELQATLHDFRQILQTRGALSLOGSIMTVGKVFSSNGQSIT 147  
QY 169 YRESLHCRIRGMLAMPKDAANTLIADYVAKSGFFRVTIGVNDLRECOYMTDNTPL 228  
DB 148 FDAIQACARAGGRIAVRNPENEAIASFVKKYNTY-AVVGTEGSPGDFRYSDGTPV 206  
QY 229 QNYNNWNEGEPDYPGHEDCVEMLSGRWNTDECHLTMFYWCFE 272  
DB 207 -NYTNWYRGEPAK-RGKQCQVEMTQGNDRNCLYSRLTICDF 248  
RESULT 14  
LNHUP6  
Pulmonary surfactant protein A precursor (clone 6A) - human  
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999  
C:Accession: A25720  
R:Floros, J.; Steinbrink, R.; Jacobs, K.; Phelps, D.; Kriz, R.; Recny, M.; Sultzman,  
J. Biol. Chem. 261, 9029-9033, 1986  
A:Title: Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfa  
A:Reference number: A25720; MUID:86250832  
A:Accession: A25720  
A:Molecule type: mRNA  
A:Residues: 1-248 <FLO>  
A:Cross-references: GB:M13686; NID:g190669; PIDN:AAA60211.1; PID:g190670  
A:Note: part of the sequence was confirmed by protein sequencing  
A:Note: the amino end of the mature protein, which was not identified, is partially a  
A:Note: clones corresponding to two different proteins were sequenced. Cotranslation  
C:Genetics:  
A:Gene: GDB:SFTPAL; SFTPL; SP-A; SP-AL  
A:Cross-references: GDB:119593; OMIM:178630  
A:Map position: 10q22-10q23  
C:Superfamily: mannose-binding lectin; C-type lectin homology  
C:Keywords: acetylated amino end; alveolar proteinosis; calcium; gaseous exchange; gl  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-248/Product: pulmonary surfactant protein A #status predicted <MAT>  
F:127-246/Domain: C-type lectin homology <LCH>  
F:21/Modified site: acetylated amino end (Glu) (in mature form) #status predicted  
F:30,33,36,42,54,57,63,76,79,82,91,97/Modified site: 4-hydroxyproline (Pro) #status p  
F:207/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 19.1%; Score 284; DB 1; Length 248;  
Best Local Similarity 29.9%; Pred. No. 3.6e-16;  
Matches 78; Conservative 42; Mismatches 119; Indels 22; Gaps 9;  
QY 16 LVLFLLQISGLDLSRPTAEVCAHTISGPGDGEKGGKGVKRMGPKGI 75  
DB 6 LALNLIILMAASGAACEVK---DVCVG---SPGIPGTPGSHGLPGRDRGLKGLDGGPPGP 59  
QY 76 KGEIGDMG-DRGNIGTKGPIKKGDKGKGLLIGPGEKAGTVCDCGRYKRFVGLDLS 134  
DB 60 MGPPGEMPCPPGNDGLPGAIFGEGCEK---GEPGPGPLPAHLDE-----ELQAT 110  
QY 135 IARLTKSMKFKVKNVIA---GIRETEEFYIVQEEKYRESLHCRIRGMLAMPKDEAA 191  
DB 111 LHDFRQILQTRGALSLOGSIMTVGKVFSSNGQSITDAIQACARAGGRIAVRNPPEE 170  
QY 192 NTLIADYVAKSGFFRVTIGVNDLRECOYMTDNTPLQNYNNWNEGEPDYPGHEDCVEM 251  
DB 171 NEAIAFVKKYNTY-AVVGTEGSPGDFRYSDGTPV-NYTNWYRGEPAK-RGKQCQVEM 227  
QY 252 LSSGRWNTDECHLTMFYWCFE 272  
DB 228 YTGQWDRNDRNCLYSRLTICDF 248

Search completed: July 3, 2002, 12:35:22  
Job time: 191 sec

```
RESULT 15
LNDGPS
pulmonary surfactant protein A precursor - dog
N:Alternate names: pulmonary surfactant 32K apoprotein; pulmonary surfactant-associated
C:Species: Canis lupus familiaris (dog)
C:date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: A25296; A61227; A60142
R:Benson, B.; Hawgood, S.; Schilling, J.; Clements, J.; Damm, D.; Cordell, B.; White, R.
Proc. Natl. Acad. Sci. U.S.A. 82, 6379-6383, 1985
A:title: Structure of canine pulmonary surfactant apoprotein: cDNA and complete amino ac
A:Reference number: A25296; MUID:86016705
A:Accession: A25296
A:Molecule type: mRNA
A:Residues: 1-248 <BEN>
A:Note: the authors translated the codon WTG for residue 60 as Pro
R:Liau, D.F.; Ryan, S.F.
Chem. Phys. Lipids 59, 29-38, 1991
A:title: Purification of surfactant protein A from dog lung by reconstitution with surfa
A:Reference number: A61227; MUID:92163993
A:Accession: A61227
A:Molecule type: protein
A:Residues: 18-32 <LIA>
R:Ross, G.F.; Meuth, J.; Ohning, B.; Kim, Y.; Whitsett, J.A.
Biochim. Biophys. Acta 870, 267-278, 1986
A:title: Purification of canine surfactant-associated glycoproteins A. Identification of
A:Reference number: A60142; MUID:86159848
A:Accession: A60142
A:Molecule type: protein
A:Residues: 24-34;95-101,'X',103-108 <ROS>
R:Pathy, L.
Nature 325, 490, 1987
A:Reference number: A93388; MUID:87115834
A:Contents: annotation; animal lectin domain homology
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t
C:Comment: This protein is a sialoglycoprotein synthesized by alveolar type II cells. It
pendent on the presence of calcium ions.
C:Superfamily: mannose-binding lectin; C-type lectin homology
F:Keywords: alveolar proteinosis; calcium; disulfide bond; gaseous exchange; hydroxyprol
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-248/Product: pulmonary surfactant protein A #status experimental <MPT>
F:28-102/Region: collagen-like
F:127-246/Domain: C-type lectin homology <LCH>
F:20,207/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:30/Modified site: 4-hydroxyproline (Pro) #status experimental

Query Match 18.9% Score 280; DB 1; Length 248;
Best Local Similarity 30.0%; Pred.No. 7.6e-16;
Matches 81; Conservative 36; Mismatches 115; Indels 38; Gaps 11;

QY 14 ILLVFLQIQSLGIDISDPTAEVCATHITSPGKGDGDEKSGEGKKGKVGKMGPK 73
: : : : : : : : : : : : : : : : : : : : : : : :
Db 6 LALALTLLWVSGI-----ENNTKDCVCG--NPGIPGTPGSHGLPGDGRGVKGDGPP 57
: : : : : : : : : : : : : : : : : : : : : : : :
QY 74 GIKGELGDM-GDRGNIGTKTGPIGKKGDKGKGLLGPCKGKAGTVCDCGRYKRVFGQLD 132
: : : : : : : : : : : : : : : : : : : : : : : :
Db 58 GPLGPGGMPGHPGNGMTGAPVAGERGEK---GEPGERGPG-----LPASLD 104
: : : : : : : : : : : : : : : : : : : : : : : :
QY 133 ISIAFLKTKMKFKVKNV---AGIRETEE-----KYYIVQEKKNYRESLTHCRIGGM 182
: : : : : : : : : : : : : : : : : : : : : : : :
Db 105 ---ELQTTIHDURHQIQLQTMGVLSLHESLLVVGKRVFSSGAOSINFNDIQELCAGAGQ 161
: : : : : : : : : : : : : : : : : : : : : : : :
QY 183 LAMPKDEAANTLIADYVAKSGFFRVFVGVNDLREGQYMFDTNPTLQNTYNNNEGEPSDP 242
: : : : : : : : : : : : : : : : : : : : : : : :
Db 162 IAAPNSPENEAVASIVKKYNTY-AYLGLVESPDSSDFQYMDGAPV-NYTNWYPCGPRG- 218
: : : : : : : : : : : : : : : : : : : : : : : :
QY 243 YGHEDCVEMLSGRWNDECHLTWYFVCEP 272
: : : : : : : : : : : : : : : : : : : : : : : :
Db 219 RGKEQCQVMTDQGNKNKNCQYRLAICEF 248
: : : : : : : : : : : : : : : : : : : : : : : :
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:28:06 ; Search time 51.7 Seconds  
(without alignments)  
595.115 Million cell updates/sec

Title: US-09-600-932-2

Perfect score: 1484

Sequence: 1 MNGFASLLRNQFILLVFL.....NDTECHLTMVCFEIKKK 277

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1484	100.0	277	20	AA1980
2	1472	99.2	277	20	AA1981
3	1472	99.2	277	21	AA1982
4	1472	99.2	277	21	AA1983
5	716.5	48.3	277	21	AA1984
6	716.5	48.3	277	21	AA1985
7	716.5	48.3	277	21	AA1986
8	716.5	48.3	277	21	AA1987
9	696.5	46.9	292	22	AA1988
10	693.5	46.7	263	22	AA1989
11	623.5	42.0	268	22	AA1990

12	376	25.3	222	22	ABG15021	Novel human diago
13	337	22.7	375	21	AA177990	Human SP-D amino a
14	328.5	22.1	351	18	AA18780	Bovine conglutinin
15	328.5	22.1	351	18	AA13672	Bovine conglutinin
16	328.5	22.1	371	16	AA175642	Bovine conglutinin
17	293.5	19.8	249	21	AA15401	Pig serum lectin.
18	291.5	19.6	248	21	AA17988	Human MBP amino ac
19	289.5	19.5	228	20	AA17369	Mannan-binding pro
20	289.5	19.5	248	14	AA15005	Human Mannose-Bind
21	289.5	19.5	248	20	AA129485	Human mannan-bind
22	288.5	19.4	248	22	AA136578	Human mannan bindi
23	288.5	19.4	248	9	AA182980	Sequence deduced f
24	287	19.3	248	21	AA177989	Human SP-A amino a
25	285.5	19.2	248	7	AA160666	Genomic sequence o
26	285.5	19.2	248	7	AA160661	Plasmid PASPc-SV(
27	285.5	19.2	271	7	AA160661	Genomic sequence o
28	284	19.1	259	21	AA158135	Lung cancer associ
29	284	19.1	335	12	AA112222	Conglutinin, Bos
30	283.5	19.1	248	7	AA160665	Sequence of human
31	283.5	19.1	248	7	AA160441	Plasmid PASPc-SV(
32	283.5	19.1	248	9	AA180694	Sequence deduced f
33	283.5	19.1	248	11	AA1804215	Human 32K ASP enco
34	283.5	19.1	248	11	AA1804216	Human 32K ASP enco
35	283.5	19.1	248	11	AA1806331	Human alveolar sur
36	283.5	19.1	271	11	AA1804212	Human 32K alveolar
37	283.5	19.1	271	11	AA1804217	Human 32K ASP enco
38	283	19.1	256	11	AA1804210	Canine 32K alveola
39	282	19.0	248	7	AA160437	Dog 32 kd alveolar
40	279	18.8	248	8	AA170662	35kd pulmonary sur
41	279	18.7	248	8	AA170663	Vector PSP 35K-1A-
42	276	18.6	248	11	AA180591	Amino acid sequenc
43	274	18.5	742	22	AA163347	Human secreted pro
44	273	18.4	457	20	AA173633	Human secreted pro
45	273	18.4	542	20	AA173634	

ALIGNMENTS

RESULT 1

AA175518  
ID AA175518 standard; Protein; 277 AA.

AC AA175518;

DT 30-SEP-1999 (first entry)

XX Human collectin protein.

XX Collectin; human; antibacterial; antiviral; treatment; infection.

XX Homo sapiens.

XX WO9937767-A1.

XX 29-JUL-1999.

XX 24-JUL-1998; 98WO-JP03328.

XX 23-JAN-1998; 98JP-0011281.

XX (FUSO ) FUSO PHARM IND LTD.

XX Wakamiya N;

XX WPI; 1999-458691/38.

XX N-PSDB; AA188323.

XX New collectin protein of human origin and DNA encoding it

XX Claim 1; Page 42-44; 58pp; Japanese.

XX This invention describes the isolation and characterisation of a novel

CC human collectin protein and its encoding polynucleotide. The human  
CC collectin exhibits antibacterial and antiviral activity and can be used  
CC as an agent for the treatment of human bacterial and viral infections.  
XX This sequence represents the novel human collectin.

SQ Sequence 277 AA;

Query Match 100.0%; Score 1484; DB 20; Length 277;  
Best/Local Similarity 100.0%; Pred. No. 4,7e-142;  
Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNGFASLRNQIFLLVLELQISGLIDSRPTAEVCATHTISPGKGDGEGDGE 60  
Db 1 mngfasllrrngfillvllqilqslgldisrptaevcathtispgkddgkdgcd 60  
QY 61 EKHGKVGMPGKIGKELGDMGDRNGTKGTPIGKKGDKGKGLLIPGKAGTVD 120  
Db 61 egkhgkvgrmpgkigkeldmgdrgnktgtpigkkgdkgkglilgpgkagktvd 120  
QY 121 CGYRKFQGLDISIARLTKSMKFKVKNVIAGIRETEKEFYIVQEKNYRESLTHCRIRG 180  
Db 121 cgyrkfvgqldisarltsmkfkvnviagireteekfyivqeknyreslthcirg 180  
QY 181 GMLAMPKDEAANTLIADYVAKSGFFRVPIGVNLEREGQYMTNTPLQYSNWNEGEPS 240  
Db 181 gmlampkdeaanliadyvaksgffrvfigvndleregymfndtntplqysnwnegeps 240  
QY 241 DPYGHEDCVEMLSGRWNTECHLPMFVCEFIKKK 277  
Db 241 dpyghedcvemlssgrwndtechltmyfvcfefikkk 277

RESULT 2

AAAY1698  
ID AAY41698 standard; Protein; 277 AA.

AC AAY41698;

XX 07-DEC-1999 (first entry)

XX Human PRO702 protein sequence.

XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
KW secreted protein; transmembrane protein.

OS Homo sapiens.

XX WO9946281-A2.

XX 16-SEP-1999.

XX 08-MAR-1999; 99WO-US05028.

XX 10-MAR-1998; 98US-0077450.

XX 11-MAR-1998; 98US-0077632.

XX 11-MAR-1998; 98US-0077641.

XX 11-MAR-1998; 98US-0077649.

XX 12-MAR-1998; 98US-0077731.

XX 13-MAR-1998; 98US-0078004.

XX 17-MAR-1998; 98US-0040220.

XX 20-MAR-1998; 98US-0078886.

XX 20-MAR-1998; 98US-0078910.

XX 20-MAR-1998; 98US-0078936.

XX 25-MAR-1998; 98US-0078939.

XX 26-MAR-1998; 98US-0079294.

XX 27-MAR-1998; 98US-0079656.

XX 27-MAR-1998; 98US-0079663.

XX 27-MAR-1998; 98US-0079664.

XX 27-MAR-1998; 98US-0079689.

XX 27-MAR-1998; 98US-0079728.

XX 27-MAR-1998; 98US-0079786.

PR 30-MAR-1998; 98US-0079920.  
PR 30-MAR-1998; 98US-0079923.  
PR 31-MAR-1998; 98US-0080105.  
PR 31-MAR-1998; 98US-0080107.  
PR 31-MAR-1998; 98US-0080165.  
PR 31-MAR-1998; 98US-0080194.  
PR 01-APR-1998; 98US-0080327.  
PR 01-APR-1998; 98US-0080328.  
PR 01-APR-1998; 98US-0080333.  
PR 01-APR-1998; 98US-0080334.  
PR 08-APR-1998; 98US-0081049.  
PR 08-APR-1998; 98US-0081070.  
PR 08-APR-1998; 98US-0081071.  
PR 09-APR-1998; 98US-0081195.  
PR 09-APR-1998; 98US-0081203.  
PR 09-APR-1998; 98US-0081229.  
PR 15-APR-1998; 98US-0081817.  
PR 15-APR-1998; 98US-0081838.  
PR 15-APR-1998; 98US-0081952.  
PR 15-APR-1998; 98US-0081955.  
PR 21-APR-1998; 98US-0082568.  
PR 21-APR-1998; 98US-0082569.  
PR 22-APR-1998; 98US-0082700.  
PR 22-APR-1998; 98US-0082704.  
PR 22-APR-1998; 98US-0082804.  
PR 23-APR-1998; 98US-0082767.  
PR 23-APR-1998; 98US-0082796.  
PR 27-APR-1998; 98US-0083336.  
PR 28-APR-1998; 98US-0083322.  
PR 29-APR-1998; 98US-0083392.  
PR 29-APR-1998; 98US-0083495.  
PR 29-APR-1998; 98US-0083496.  
PR 29-APR-1998; 98US-0083499.  
PR 29-APR-1998; 98US-0083500.  
PR 29-APR-1998; 98US-0083545.  
PR 29-APR-1998; 98US-0083554.  
PR 29-APR-1998; 98US-0083558.  
PR 29-APR-1998; 98US-0083559.  
PR 30-APR-1998; 98US-0083742.  
PR 05-MAY-1998; 98US-0084366.  
PR 06-MAY-1998; 98US-0084414.  
PR 06-MAY-1998; 98US-0084441.  
PR 07-MAY-1998; 98US-0084598.  
PR 07-MAY-1998; 98US-0084600.  
PR 07-MAY-1998; 98US-0084637.  
PR 07-MAY-1998; 98US-0084639.  
PR 07-MAY-1998; 98US-0084640.  
PR 07-MAY-1998; 98US-0084643.  
PR 13-MAY-1998; 98US-0085323.  
PR 13-MAY-1998; 98US-0085338.  
PR 13-MAY-1998; 98US-0085339.  
PR 15-MAY-1998; 98US-0085573.  
PR 15-MAY-1998; 98US-0085579.  
PR 15-MAY-1998; 98US-0085580.  
PR 15-MAY-1998; 98US-0085582.  
PR 15-MAY-1998; 98US-0085689.  
PR 15-MAY-1998; 98US-0085697.  
PR 15-MAY-1998; 98US-0085700.  
PR 15-MAY-1998; 98US-0085704.  
PR 18-MAY-1998; 98US-0086023.  
PR 22-MAY-1998; 98US-0086392.  
PR 22-MAY-1998; 98US-0086414.  
PR 22-MAY-1998; 98US-0086430.  
PR 22-MAY-1998; 98US-0086486.  
PR 28-MAY-1998; 98US-0087098.  
PR 28-MAY-1998; 98US-0087106.  
PR 30-JUL-1998; 98US-0087208.  
PR 11-SEP-1998; 98US-0094651.  
PR 11-SEP-1998; 98US-0100038.

(GETH ) GENENTECH INC.

XX PA

XX XX



PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
 XX WPI; 1999-551358/46.  
 DR N-PSDB; AAZ33973.  
 XX  
 PT New secreted and transmembrane polypeptides and their polynucleotides,  
 PT useful for treating blood coagulation disorders, cancers and cellular  
 PT adhesion disorders -  
 XX  
 PS Claim 12; Flg 37; 530pp; English.  
 XX  
 CC The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC sources of probes, primers, for chromosome mapping, and for generation  
 CC of antisense sequences. They can also be used to create transgenic  
 CC animals. The proteins can be used to treat a variety of diseases and  
 CC disorders, depending on their function. Diseases that may be treated  
 CC include blood coagulation disorders, cancers and cellular adhesion  
 CC disorders. They may also be used to raise antibodies. AAZ33891 to  
 CC AAZ34338, and AAZ41685 to AAZ41774 represent polynucleotide and  
 CC polypeptide sequence given in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 277 AA;

Query Match 99.2%; Score 1472; DB 20; Length 277;  
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;  
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MNGFASLLRNQFILLVFLQISGLDIDSRPTAEVCATHTISPGKGDGKGDGPE 60  
 Db 1 mngfasllrrnqfllvflqlqsgldidsrptaevcathtispkgddgkdgpe 60  
 QY 61 EGRHGKVGMPGKIGKELGDMGDRGNIGKTGPIGKKGDKGKGLIGIPGKAGTVC 120  
 Db 61 egkhgkvgrmpgkigkeldmgdrgnigktgpihgkkgdkgkgligipgkagtvcd 120  
 QY 121 CGRYRKFVGQDISIARLKTSMKFVKNVIAGIRETEKFFYIVQEKNYRESLTHCRIRG 180  
 Db 121 cgyrkfvgqdisiarlktsmkfvknviagireteekffyyivqeknyreslthcrirg 180  
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRFGVNDLEREGQVMTDNTPLQNSWNNEGEP 240  
 Db 181 gmlampkdeantliadyvaksffrfgvndlerregqvmstdntplqnswnnegeps 240  
 QY 241 DPYGHEDCVMSSGRWNTDTECHLTMYFVCFEIKKKK 277  
 Db 241 dpyghedcvmssgrwndtechlmyfvcfeikkkk 277

RESULT 3  
 AAB44254  
 ID AAB44254 standard; Protein; 277 AA.  
 XX  
 AC AAB44254;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human PRO702 (UNQ366) protein sequence SEQ ID NO:97.  
 XX  
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;  
 KW expressed sequence tag; detection; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO2000053756-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 18-FEB-2000; 2000WO-US04341.  
 XX  
 PR 08-MAR-1999; 99WO-US05028.

12-MAR-1999; 99US-0123957.  
 PR 29-MAR-1999; 99US-0126773.  
 PR 21-APR-1999; 99US-0130232.  
 PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US283113.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;  
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Kijavini IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;  
 XX  
 XX WPI; 2000-611443/58.  
 DR N-PSDB; AAZ78480.  
 DR  
 XX

Novel PRO polypeptides and polynucleotides used in detection methods,  
 PT to target bioactive molecules to specific cells, and to modulate  
 PT cellular activities -  
 XX  
 PS Claim 12; Fig 37; 636pp; English.  
 XX  
 CC AAZ78458 to AAZ78599 represent polynucleotide and EST (expressed  
 CC sequence tag) sequences which encode secreted or transmembrane PRO  
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
 CC activity. The polynucleotides and polypeptides can be used for detecting  
 CC the presence of PRO polypeptides in samples, for linking bioactive  
 CC molecules to cells and for modulating biological activities of cells,  
 CC using the polypeptides for specific targeting. The polypeptide targeting  
 CC can be used to kill the target cells, e.g. for the treatment of cancers.  
 CC The polypeptide pairs provide specific targeting of bioactive molecules  
 CC to cells. AAZ78600 to AAZ78987 represent PCR primers and probes used in  
 CC the isolation of the PRO polynucleotide sequences.  
 XX  
 SQ Sequence 277 AA;

Query Match 99.2%; Score 1472; DB 21; Length 277;  
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;  
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MNGFASLLRNQFILLVFLQISGLDIDSRPTAEVCATHTISPGKGDGKGDGPE 60  
 Db 1 mngfasllrrnqfllvflqlqsgldidsrptaevcathtispkgddgkdgpe 60  
 QY 61 EGRHGKVGMPGKIGKELGDMGDRGNIGKTGPIGKKGDKGKGLIGIPGKAGTVC 120  
 Db 61 egkhgkvgrmpgkigkeldmgdrgnigktgpihgkkgdkgkgligipgkagtvcd 120  
 QY 121 CGRYRKFVGQDISIARLKTSMKFVKNVIAGIRETEKFFYIVQEKNYRESLTHCRIRG 180  
 Db 121 cgyrkfvgqdisiarlktsmkfvknviagireteekffyyivqeknyreslthcrirg 180  
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRFGVNDLEREGQVMTDNTPLQNSWNNEGEP 240  
 Db 181 gmlampkdeantliadyvaksffrfgvndlerregqvmstdntplqnswnnegeps 240  
 QY 241 DPYGHEDCVMSSGRWNTDTECHLTMYFVCFEIKKKK 277  
 Db 241 dpyghedcvmssgrwndtechlmyfvcfeikkkk 277

RESULT 4  
AAU29073  
ID AAU29073 standard; Protein; 277 AA.  
XX AC AAU29073;  
XX DT 18-DEC-2001 (first entry)  
XX DE Human PRO polypeptide sequence #50.  
XX KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;  
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200168848-A2.  
XX PD 20-SEP-2001.  
XX PE 28-FEB-2001; 2001WO-US06520.  
XX PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 03-MAR-2000; 2000US-187202P.  
PR 06-MAR-2000; 2000US-186968P.  
PR 14-MAR-2000; 2000US-189320P.  
PR 14-MAR-2000; 2000US-189328P.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 21-MAR-2000; 2000US-190828P.  
PR 21-MAR-2000; 2000US-191007P.  
PR 21-MAR-2000; 2000US-191048P.  
PR 21-MAR-2000; 2000US-191314P.  
PR 26-MAR-2000; 2000US-192655P.  
PR 29-MAR-2000; 2000US-193032P.  
PR 29-MAR-2000; 2000US-193053P.  
PR 30-MAR-2000; 2000WO-US08439.  
PR 04-APR-2000; 2000US-194449P.  
PR 04-APR-2000; 2000US-194677P.  
PR 11-APR-2000; 2000US-195975P.  
PR 11-APR-2000; 2000US-196000P.  
PR 11-APR-2000; 2000US-196187P.  
PR 11-APR-2000; 2000US-196690P.  
PR 11-APR-2000; 2000US-196820P.  
PR 18-APR-2000; 2000US-198121P.  
PR 18-APR-2000; 2000US-198585P.  
PR 25-APR-2000; 2000US-199397P.  
PR 25-APR-2000; 2000US-199550P.  
PR 25-APR-2000; 2000US-199654P.  
PR 03-MAY-2000; 2000US-201516P.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 05-JUN-2000; 2000US-209832P.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 22-AUG-2000; 2000US-0644848.  
PR 24-AUG-2000; 2000US-0644848.  
PR 08-NOV-2000; 2000WO-US33328.  
PR 01-DEC-2000; 2000WO-US30952.  
PR 20-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
XX PA (GETH ) GENENTECH INC.  
XX PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PU, Gurney AL;  
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
XX WPI; 2001-602746/68.  
XX DR N-PSDB; AAS45974.  
XX

PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
PT presence of tumours, such as prostate and breast tumours, in mammals and  
XX to screen for modulators of the compounds -  
XX Claim 11; Fig 100; 774pp; English.  
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.  
CC The PRO polypeptides and their associated nucleic acids can be used to  
CC detect the presence of a tumour in a mammal by comparing the level of  
CC expression of a PRO polypeptide in a test sample of cells from the animal  
CC and a control sample of normal cells, whereby a higher level of  
CC expression in the test sample indicates the presence of a tumour in the  
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats  
CC and rabbits but are preferably human. The polypeptides can be used to  
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,  
CC when contacted with it. A specific polypeptide can be used to stimulate  
CC the proliferation or differentiation of chondrocyte cells. The PRO  
CC proteins can be used to determine the presence of tumours and also  
CC susceptibility to tumour development, particularly adrenal, lung, colon,  
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
CC can be used for genetic analysis of individuals with genetic disorders.  
XX Sequence 277 AA;  
SQ

Query Match 99.2%; Score 1472; DB 22; Length 277;  
Best Local Similarity 99.3%; Pred. No. 7.8e-141;  
Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MNGFASLLRRNQFILLVFLQLQSLGLDIDSRPTAEVCATHTISPGKGGDGEKGPGE 60  
DB 1 mngfasllrrnqfllvflqlqslgldidsrptaevcathtispgkggdgekdpge 60  
QY 61 BGKHGKVGWGPCKIGKELGMDGRNICKTGPKCKYKGDGKGLLGPBGKAGTVCD 120  
DB 61 egkhgkvgrmgpckigkeldgmdgdnigktgpigkkgdkgkglgipgekagtvcd 120  
QY 121 CGRYKFEVGLDIIARLKTSMFVKVNIAGIRTEEEKFYIIVQEEKYRESLTHCRIRG 180  
DB 121 cgrykfvglldiisarlktsmfkvkvnigagireteekfyiivqeeekyreslthcrrg 180  
QY 181 GMLAMPKDEAANTLIADYVAKSGFFRVFIGVNDLEREGQYMETDNTPLQNSWNNEGEP 240  
DB 181 gmlampkdeaanltiadyvaksgffrvfigvndlerregqymstdntplqnswnnegeps 240  
QY 241 DPYGHEDCVEMLSGRWNTDECHLTMYFVCFEIRKKK 277  
DB 241 dpyghedcvemlssgrwnndtechltmyfvcfeikkk 277

RESULT 5  
AAB24075  
ID AAB24075 standard; Protein; 271 AA.  
XX AC AAB24075;  
XX DT 29-JAN-2001 (first entry)  
XX DE Human PRO1182 protein sequence SEQ ID NO:51.  
XX KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;  
KW proliferation; tumorigenesis; identification; cancer; cytostatic;  
KW neurotropic; neuroprotective; antiinflammatory; immunosuppressive;  
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;  
KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;  
KW hypothalamic disorder; glandular disorder; macrophagal disorder;  
KW epithelial disorder; stromal disorder; blastocoeleic disorder;  
XX inflammatory disorder; immunologic disorder.  
XX OS Homo sapiens.  
XX PN WO2000053755-A2.



PR	09-JUN-1998	9805-00865538
PR	10-JUN-1998	9805-00887222
PR	10-JUN-1998	9805-00887300
PR	10-JUN-1998	9805-00887334
PR	10-JUN-1998	9805-00887388
PR	10-JUN-1998	9805-00887400
PR	10-JUN-1998	9805-00887410
PR	10-JUN-1998	9805-00887422
PR	10-JUN-1998	9805-00888100
PR	10-JUN-1998	9805-00888110
PR	10-JUN-1998	9805-00888224
PR	10-JUN-1998	9805-00888258
PR	11-JUN-1998	9805-00888266
PR	11-JUN-1998	9805-00888588
PR	11-JUN-1998	9805-00888611
PR	11-JUN-1998	9805-00888636
PR	11-JUN-1998	9805-00888766
PR	12-JUN-1998	9805-00890390
PR	12-JUN-1998	9805-00891010
PR	16-JUN-1998	9805-00894440
PR	16-JUN-1998	9805-00894512
PR	16-JUN-1998	9805-00895114
PR	17-JUN-1998	9805-00895338
PR	17-JUN-1998	9805-00895382
PR	17-JUN-1998	9805-00895598
PR	17-JUN-1998	9805-00895599
PR	17-JUN-1998	9805-00896000
PR	18-JUN-1998	9805-00896533
PR	18-JUN-1998	9805-00898011
PR	18-JUN-1998	9805-00899007
PR	19-JUN-1998	9805-00899508
PR	19-JUN-1998	9805-00899477
PR	19-JUN-1998	9805-00899548
PR	22-JUN-1998	9805-00899552
PR	22-JUN-1998	9805-00902466
PR	22-JUN-1998	9805-00902552
PR	23-JUN-1998	9805-00902534
PR	23-JUN-1998	9805-00903449
PR	23-JUN-1998	9805-00903555
PR	24-JUN-1998	9805-00904029
PR	24-JUN-1998	9805-00904131
PR	24-JUN-1998	9805-00904335
PR	24-JUN-1998	9805-00904444
PR	24-JUN-1998	9805-00904455
PR	24-JUN-1998	9805-00904611
PR	24-JUN-1998	9805-00904772
PR	24-JUN-1998	9805-00905335
PR	24-JUN-1998	9805-00905340
PR	24-JUN-1998	9805-00905548
PR	24-JUN-1998	9805-00905557
PR	25-JUN-1998	9805-00906676
PR	25-JUN-1998	9805-00906788
PR	25-JUN-1998	9805-00906868
PR	25-JUN-1998	9805-00906930
PR	25-JUN-1998	9805-00906911
PR	25-JUN-1998	9805-00906944
PR	25-JUN-1998	9805-00906955
PR	26-JUN-1998	9805-00906936
PR	26-JUN-1998	9805-00906862
PR	01-JUL-1998	9805-00913358
PR	01-JUL-1998	9805-00913600
PR	01-JUL-1998	9805-00915544
PR	02-JUL-1998	9805-00914778
PR	02-JUL-1998	9805-00914866
PR	02-JUL-1998	9805-00915119
PR	02-JUL-1998	9805-00916236
PR	02-JUL-1998	9805-00916298
PR	02-JUL-1998	9805-00916333
PR	02-JUL-1998	9805-00916446
PR	02-JUL-1998	9805-00916773
PR	07-JUL-1998	9805-00919812
PR	07-JUL-1998	9805-00919828

XX	Homo sapiens.	
OS		
XX	WO200073454-A1.	
PN		
XX	07-DEC-2000.	
PD		
PD		
XX		
XX	30-MAR-2000; 2000WO-US08439.	
PF		
XX		
XX	02-JUN-1999; 99WO-US12252.	
PR	23-JUN-1999; 99US-0141037.	
PR	07-JUL-1999; 99US-0143048.	
PR	20-JUL-1999; 99US-0144758.	
PR	26-JUL-1999; 99US-0145698.	
PR	28-JUL-1999; 99US-0146222.	
PR	17-AUG-1999; 99US-0149396.	
PR	15-SEP-1999; 99WO-US21090.	
PR	15-SEP-1999; 99WO-US21547.	
PR	08-OCT-1999; 99US-0158663.	
PR	30-NOV-1999; 99WO-US28313.	
PR	01-DEC-1999; 99WO-US28301.	
PR	16-DEC-1999; 99WO-US30095.	
PR	20-DEC-1999; 99WO-US30911.	
PR	05-JAN-2000; 2000WO-US00219.	
PR	06-JAN-2000; 2000WO-US00376.	
PR	11-FEB-2000; 2000WO-US03565.	
PR	18-FEB-2000; 2000WO-US04341.	
PR	22-FEB-2000; 2000WO-US04414.	
PR	24-FEB-2000; 2000WO-US04914.	

XX CC The invention provides membrane-bound PRO polypeptides and  
CC polynucleotides encoding them. The PRO sequences of the invention were  
CC identified based on extracellular domain homology screening. The PRO  
CC sequences have homology with proteins including LDL receptors, TIE  
CC ligands and various enzymes. The membrane-bound proteins and receptor  
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor  
CC immunoadhesins, for instance, can be used as therapeutic agents to block  
CC receptor-ligand interactions. The membrane-bound proteins can also be  
CC employed for screening of potential peptide or small molecule inhibitors  
CC of the relevant receptor/ligand interaction. The PRO encoding sequences  
CC are useful as hybridization probes, in chromosome and gene mapping and in  
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences

PR 24-FEB-2000; 2000WO-US05004.  
 PR 02-MAR-2000; 2000WO-US05841.  
 PR 15-MAR-2000; 2000WO-US06884.  
 PR 20-MAR-2000; 2000WO-US07377.  
 XX (GETH ) GENENTECH INC.  
 XX  
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;  
 PI Grimaldi CJ, Gurney AL, Kljavin LJ, Napier MA, Pan J, Paoni NF;  
 PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;  
 PI Zhang Z;  
 XX  
 XX WPI: 2001-032160/04.  
 DR N-PSDB; AAF44230.  
 XX  
 XX PRO polynucleotides used to produce polypeptides used to target  
 PT bioactive molecules such as toxins, radiolabels or antibodies, to  
 PT specific cells, to cause targeted cell death -  
 XX  
 PS Claim 12; Fig 252; 935pp: English.  
 XX  
 CC The present invention describes human secreted and transmembrane PRO  
 CC proteins. The PRO proteins have cytosolic activity. The PRO proteins  
 CC can be used for targeted delivery of bioactive molecules, such as  
 CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide  
 CC sequences and their fragments, can be used as hybridisation probes, in  
 CC chromosomal and gene mapping, and in the generation of anti-sense RNA  
 CC and DNA. They may also be used to produce transgenic animals which are  
 CC used to develop and screen therapeutically useful reagents. The PRO  
 CC nucleotide and protein sequence can be used for tissue typing and in  
 CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.  
 CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and  
 CC AAF5154 to AAF5300 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.  
 XX  
 SQ Sequence 271 AA;  
 XX  
 Query Match 48.3%; Score 716.5; DB 22; Length 271;  
 Best Local Similarity 47.4%; Pred. No. 3e-64;  
 Matches 129; Conservative 60; Mismatches 76; Indels 7; Gaps 3;  
 QY 8 LERNQFLLVLLFLLQSLGLDIDSRPTA-EVCATHITISPGKDGKDGKDGKDGK 66  
 Db 1 mrgnlavglvllslflpsghpdpagddacsqvlvpglkgdagekgd---kgapgr 57  
 QY 67 VGRMGPKGKIGELGMDGRGNI---GKTGPIGKKGDKGKGLLGPGEKGRAGTVCDCGR 123  
 Db 58 puryvgtgkdgmgkgkgvgrhkgipgskgkdgsgdipppngpgeplpccsq 117  
 QY 124 YKFKVGOLDISTARLKTSMKFKVKNVIAGIRETEEFYIVQEEKNYRESLTHCRIRGGM 183  
 Db 118 lrfalgmdqvsqsltselkfnknagvreteskiylvkeekyadaqlscgrggtl 177  
 QY 184 AMPKDEANTLIADYVAKSGFFRVFGVNDLEREGQYMTDNTPLQNSNNNEGPD 243  
 Db 178 smpkdeanqlmaaylagaglarvfigindlekafvysdhspmrtnfkwsrsgpunay 237  
 QY 244 GHEDCVMLSSGWNNDTECHTMYFVCEFIKK 275  
 Db 238 deedcvmvasggnvachttmyfmcfdke 269  
 RESULT 9  
 AAM78426  
 ID AAM78426 standard; Protein; 292 AA.  
 XX  
 AC AAM78426;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX

DE Human protein SEQ ID NO 1088.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 XX 05-FEB-2001; 2001WO-US04098.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue A, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 XX WPI: 2001-476283/51.  
 DR N-PSDB; AAK51559.  
 XX  
 XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 PS Claim 20; Page 3318-3319; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation of which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 XX  
 SQ Sequence 292 AA;  
 XX  
 Query Match 46.9%; Score 696.5; DB 22; Length 292;  
 Best Local Similarity 46.2%; Pred. No. 3.6e-62;  
 Matches 126; Conservative 58; Mismatches 80; Indels 9; Gaps 3;  
 QY 3 GFASLLRRNQFLLVLLFLLQSLGLDIDSRPTA---EVCATHITISPGKDGKDGKDPG 59  
 Db 23 gvlpalrmrgnlavglvllslflpsghpdpagddacsqvlvpglkgdagekgd-- 80  
 QY 60 EECKHGKVGMRGMPKIGKELGMDGRGNI---GKTGPIGKKGDKGKGLLGPGEKGRAG 116  
 Db 81 -kgapgrprvgtgkdgmgkgkgvgrhkgipgskgkdgsgdipppngpgepg 139  
 QY 117 TVDCGGRYKFEVQGLDISTARLKTSMKFKVKNVIAGIRETEEFYIVQEEKNYRESLTHC 176  
 Db 140 lpccsqqlrkaigmdhngvsqsltselkfnknagvreteskiylvkeekyadaqlsc 199  
 QY 177 RINGMLAMPKDEANTLIADYVAKSGFFRVFGVNDLEREGQYMTDNTPLQNSNNNE 236



```

XX
SQ Sequence 268 AA;

Query Match 42.0%; Score 623.5; DB 22; Length 268;
Best Local Similarity 42.2%; Pred. No. 8e-55;
Matches 114; Conservative 53; Mismatches 76; Indels 27; Gaps 2;

Qy 3 GFASLLRRNQFLLVFLFLQISGLIDISRPFTA--EVCATHTISPFGKGDGKGDPG 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 23 gwlpalmrgnialvgvllslafslpsghpcpagddacsvqlvpgikdgakekdkg 82
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 60 EEGXNGVGMGPKGIKGELGDMGRGNIGKTGPICKKDGKGEGLIGIPGEGKAGTVC 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83 apgrpgvr-----gptgekgekgdsdgldgppgngpegqlpc 118
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 120 DCGYRRYFGQLDTSIARLKTSMKFKVNIAGIRETEEFYIIVOEKNYRESLTHCRIR 179
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 119 ecsqlrkaigendqvqlcseikfknavagvreteskiylvkeekryadaqlscqr 178
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 180 GGMLAMPKDEAAANTLIADYVAKSGFRRVFCVNDLEREGQYMTDNTPLQNYSNWNGEP 239
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 179 ggtlsmkpdeaanglmaaylaqaglarfigindlekegafvysdshspmtfknwrsgep 238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 240 SDPYGHEDCVEMLSGRWNTDCHLTWYFV 269
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 239 nmaydeedcvemvasgwndvachttmyfm 268
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
ABGL5021
ID ABGL5021 standard; Protein; 222 AA.
XX
AC ABGL5021;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15012.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
DR WPI; 2001-639362/73.
DR N-PSDB; AAQ79208.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID No 45380; 103pp; English.
CC
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques

```



XX	Sequence	375 AA:
5Q		

Query Match	22.1%;	Score 328.5;	DB 18;	Length 351;
Best Local Similarity	29.2%;	Pred. NO. 9.4e-25;		
Matches	84;	Conservative 33;	Mismatches 98;	Indels 73; Gaps 9;

QY	46	POP	KDGDGEKGP	-----GEGKHKGKVGEMPKIGELGDMGDRGNIGK-----TGP	93
Db	76	ppp	kgdtgprppmpg	ppagregpsakqsmgppgtppkgctgpgkvgagp.iqgfpqg	135
QY	94	ICK	KDGKEKGLIGIPG	EKGKAG-----	116
Db	136	sgl	kgkragpctgag	ragvtpsaipgppgsaqgppgkqgdgdpgetgasgesgl	195
QY	117	-----	TVCDCGRYKPVGQLDIS	IAULKTSMPFVNVTAGTRETKEFYIYVOEE	166
Db	196	aeval	kqrvtild-ghlrrfq	----afsygkavilpdpqgavg-----ekikftagav	245
QY	167	KNY	RESLTHCRIRGGMAMP	KDEAANTLIADYVAKSGFRRVFIGVNDLREGQVMETDNT	226
Db	246	ksy	aeaelcreekqgl	aspraaneavtqmv-raqeknayismndistegrftyptgc	304
QY	227	PLQWY	SNWNEGP	--SDPYGHBCDCEVMSLSSGRWNTDECHITMYFVCEF	272
Db	305	liiv	-ysnaadgennsd	eqqpcnvelfpdkwndpcskallivcef	351

RESULT 15  
AAW13672  
ID AAW13672 standard: protein; 351 AA;

OS Bos taurus.  
XX  
PN WO9707133-A

PF	02-OCT-1995;	95WO-JP02035
XX		
PR	17-AUG-1995;	95JP-0209698
XX		
PA	(FUSO ) FUSO PHARM IND LTD	

PI Wakamiya N;  
XX  
DR WPI: 1997-165243/15.

PT Recombinant conglutinin with truncated collagen region - has virus  
PT neutralising activity and can be used as an antiviral drug  
XX  
XX Disclosure: Pages 18-19; 37pp; Japanese.

CC The present sequence is bovine conglutinin (bCG) from which a

novel recombinant pCG can be prepared. The recombinant pCG comprises a truncated collagen region of formula Gly-Xaa-Xaa-Gly-Xaa-Xaa (where Xaa is any amino acid) and the neck



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:34:52 ; Search time 15.79 Seconds  
(without alignments)  
679.247 Million cell updates/sec

Title: US-09-600-932-2

Perfect score: 1484

Sequence: 1 MNGFASLLRRNQFLLEVL.....NDTECHLWYFVCFIKKKK 277

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	349.5	23.6	369	1	PSPD_BOVIN
2	344	23.2	301	1	CL43_BOVIN
3	340	22.9	374	1	PSPD_MOUSE
4	337	22.7	375	1	PSPD_HUMAN
5	333	22.4	374	1	PSPD_RAT
6	328.5	22.1	371	1	CONG_BOVIN
7	305.5	20.6	247	1	PSPA_RABIT
8	290	19.5	238	1	MABA_RAT
9	290	19.5	244	1	MABC_MOUSE
10	289.5	19.5	248	1	MABC_HUMAN
11	285.5	19.2	249	1	PSPA_PIG
12	284.5	19.2	249	1	MABC_BOVIN
13	284	19.1	248	1	PSPA_HUMAN
14	280	18.9	248	1	PSPA_CANFA
15	279	18.8	244	1	MABC_RAT
16	277	18.7	248	1	PSPA_RAT
17	270.5	18.2	239	1	MABA_MOUSE
18	270	18.2	247	1	PSPA_CAVPO
19	269.5	18.2	248	1	PSPA_MOUSE
20	207	13.9	689	1	CA29_HUMAN
21	196	13.2	1775	1	CA14_DROME
22	192.5	13.0	2944	1	CA17_HUMAN
23	191.5	12.9	1366	1	CA21_HUMAN
24	191	12.9	1022	1	CA26_CHICK
25	189.5	12.8	1018	1	CA26_HUMAN
26	187.5	12.6	518	1	MTCO_MOUSE
27	187	12.6	1029	1	CA26_MOUSE
28	186	12.5	1025	1	CA26_MOUSE
29	186	12.5	1650	1	CA2B_MOUSE
30	185.5	12.5	674	1	CA1A_CHICK
31	184.5	12.4	1362	1	CA21_CHICK
32	183.5	12.4	1364	1	CA21_BOVIN
33	183	12.3	1838	1	CA15_HUMAN

34 182 12.3 1262 1 CA13\_CHICK  
35 181.5 12.2 1019 1 CA16\_CHICK  
36 181.5 12.2 1603 1 CA1F\_HUMAN  
37 180.5 12.2 1366 1 CA21\_HUMAN  
38 180.5 12.2 1496 1 CA25\_HUMAN  
39 179.5 12.1 1028 1 CA16\_HUMAN  
40 179 12.1 520 1 MRCO\_HUMAN  
41 178.5 12.0 1372 1 CA21\_MOUSE  
42 178.5 12.0 1669 1 CA14\_HUMAN  
43 177.5 12.0 483 1 MRCO\_MESAU  
44 177.5 12.0 1418 1 CA12\_HUMAN  
45 177 11.9 547 1 CA1F\_EPHMU

P12105 gallus gall  
P20785 gallus gall  
Q07092 homo sapien  
O46392 canis famil  
P05997 homo sapien  
P12109 homo sapien  
Q9uew3 homo sapien  
Q01149 mus musculu  
P02462 homo sapien  
Q9wub9 mesocricetu  
P02458 homo sapien  
P18856 ephydatia m

## ALIGNMENTS

RESULT 1  
PSPD\_BOVIN  
ID PSPD\_BOVIN STANDARD; PRT: 369 AA.  
AC P35246;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).  
GN SFTPD OR SFTPA.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.  
RC TISSUE=Lung;  
RX MEDLINE=93170856; PubMed=8436402;  
RA Lim B.L., Lu J., Reid K.B.M.;  
RT surfactant protein D and demonstration of liver as a site of  
RT synthesis of conglutinin.  
RL Immunology 78:153-165(1993).  
CC -!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED  
MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER  
EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE  
EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.  
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: X75911; CAA53510.1;  
CC PIR: S33603; S33603.  
CC HSP: P35247; 1808.  
CC InterPro: IPR000087; Collagen.  
CC InterPro: IPR001304; lectin\_c.  
CC Pfam: PF01391; Collagen; 3.  
CC Pfam: PF00059; lectin\_c; 1.  
CC SMART: SM00034; CLECT; 1.  
CC PROSITE: PS00615; C-TYPE\_LECTIN\_1; 1.  
CC PROSITE: PS50041; C-TYPE\_LECTIN\_2; 1.  
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;  
Signal; Lectin; Collagen; Repeat; Coiled coil.

```

RESULT      2
CL43_BOVIN
ID          CL43_BOVIN          STANDARD;          PRT;          301 AA.
AC          P42916;
DT          01-NOV-1995 (Rel. 32, Created)
DT          01-NOV-1995 (Rel. 32, Last sequence update)
DT          15-JUL-1998 (Rel. 36, Last annotation update)
DE          Collectin-43 (CL-43).
DE          Bos taurus (Bovine).
OC          OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC          Bovidae; Bovinae; Bos.
OX          NCBI_TaxId=9913;
RN          [1]
RP          SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RR          TISSUE=Liver;
RX          MEDLINE=9416283; PubMed=8163480;
RX          Lim B.-L., Willis A.C., Reid K.B.M., Lu J., Laursen S.B.,
RA          Jensenius J.C., Holmskov U.:
RA          "Primary structure of bovine collectin-43 (CL-43). Comparison with
RT          conglutinin and lung surfactant protein-D." ;
RL          J. Biol. Chem. 269:11820-11824(1994).
RC          -!- FUNCTION: LECTIN THAT BINDS TO VARIOUS SUGARS: MANNOSE >
CC          FUCOSE > GLCNAC > GLUCOSE = MALTOSE > GALACTOSE > LACTOSE >
CC          GALNAC. COULD PLAY A ROLE IN IMMUNE DEFENSE.
CC          -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC          -!- SUBCELLULAR LOCATION: Extracellular.
CC          -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC          -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X75912; CAA53511.1; ALT_SEQ.
DR HSSP; P35247; 1B08.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR Lentin; Hydroxylation; Glycoprotein; Mannose-binding; Collagen;
KW Repeat; Calcium.
FT DOMAIN 29 142 COLLAGEN-LIKE.
FT FT01 202 301 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 204 299 BY SIMILARITY.
FT DISULFID 277 291 BY SIMILARITY.
SQ SEQUENCE 301 AA; 31362 MW; 0385C10B8424CD76 CRC64;
-----
Query Match 23.2%; Score 344; DB 1; Length 301;
Best Local Similarity 35.0%; Pred. No. 3e-22;
Matches 86; Conservative 33; Mismatches 99; Indels 28; Gaps 8;
-----
QY 45 SPGPKGDGEKGDPEE--GKHGKVRMGPKIGKELGMDGRGNICKTGPICKGKDGK 101
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66 SMGPPTGPGGEPGGVGAPGPGSPAGLKGKRGAPGGGAI GPQGPSAMGPPG 125
QY 102 EKGLLGIPECKKAG--TYVCDGRYRFVQGLDISIAKLTKSKFKVKNVIAGIRE----- 154
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 LKSGRDGPGKGGARGETSIVLEVDLQRQRMENLEGEVQRL-----QNIYTVKRAVLFP 178
QY 155 -----TEEFYIVIOEKNTRESLTHCRIGRMPLMKDPAANTLIADYV-AKSGFFRFVI 209
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 DGQAVGEKIFKTAGAKYSVDAEQLCREAKGQLASPRSSSENAYTLVRAKKH--HAYL 236
QY 210 GYVNDLRGGQTFMTNDPIQNTYKNWNGEESD---PYGHEDCVEMLSGRWNDETECHLT 266
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 SMNDISKCKGKTYPTGSSL-DYSNWAEPGPNRAKDEGPNCLCITYSGDNWNDECREER 295
QY 267 YFVCEG 272
Db |||
296 LVICEF 301
-----
RESULT 3
PSPD_MOUSE ID PSPD_MOUSE STANDARD; PRT; 374 AA.
AC P50404;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GS SFTPD OR SFTP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBA; TISSUE=Lung;
RX MEDLINE=96094460; PubMed=7499852;
RA Motwani M., White R.A., Guo N., Dowler L.L., Tauber A.I., Sastry K.N.;
RT "Mouse surfactant protein-D. cDNA cloning, characterization, and gene
RT localization to chromosome 14."
RL J. Immunol. 155:5671-5677 (1995).
RC -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED

```

CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER  
 CC EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE  
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.  
 CC -|- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.  
 CC -|- SUBCELLULAR LOCATION: Extracellular.  
 CC -|- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
 CC -|- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: L40156; AAA92021.1; -;  
 CC HSP: P35247; I808.  
 CC MGD; MGI:109515; Sftpd.  
 CC InterPro: IPR000087; Collagen.  
 CC InterPro: IPR001304; lectin\_c.  
 CC Pfam: PF01391; Collagen; 3.  
 CC Pfam: PF00059; lectin\_c; 1.  
 CC SMART: SM00034; CLECT; 1.  
 CC PROSITE: PS00615; C-TYPE-LECTIN.1; 1.  
 CC PROSITE: PS00041; C-TYPE-LECTIN.2; 1.  
 CC GlycoProtDB: Calcium; Surface film; Gaseous exchange; Hydroxylation;  
 KW Signal; Lectin; Collagen; Repeat; Coiled coil.  
 FT SIGNAL 1 19  
 FT CHAIN 20 374 PULMONARY SURFACTANT-ASSOCIATED PROTEIN  
 FT D.  
 FT DOMAIN 45 221  
 FT DOMAIN 222 253 COLLAGEN-LIKE.  
 FT DOMAIN 278 374 COILED COIL (POTENTIAL).  
 FT DISULFID 280 372 C-TYPE LECTIN (SHORT FORM).  
 FT DISULFID 350 364 BY SIMILARITY.  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 374 AA; 37688 MW; FE03426123F343E4 CRC64;  
 Query Match 22.9%; Score 340; DB 1; Length 374;  
 Best Local Similarity 32.4%; Pred. No. 8.5e-22;  
 Matches 80; Conservative 34; Mismatches 107; Indels 26; Gaps 6;  
 QY 46 PGPKGDDGKGD---PGEGKHGKVGKMGKIGKELGMDGRNIGTKTGPIGKKGDKGE 102  
 DB 134 PGPAGGKPGVEGAPGQGGTGAKGTGKGERGAPGVQAGPAGNAGAGAPGAPGQA 193  
 QY 103 KGLIGIPGKAGTVCDCGRYKRFVGLDI-SIARLKTSMKFKVKNVIAGI----- 152  
 DB 194 PGRGPPGLKGRGVPGD---RGIKGESGLPDSALRQMEALKGLQRLVAFSHYQK 249  
 QY 153 -----RETEKFFIYVQEKYRSLTHCRIRGMLAMPKDEANLTIADYVAKSGFF 205  
 DB 250 AALFPDGRSVGDKIFRTADSEKPFEDAEQWCKAGQGLASPRGATENAIOQLITAHN-K 308  
 QY 206 RVFIGVNDLERGQWYMTDNTPLQNTSNNEGERSDPYGHEDCVEMLSGSGWNTTECHLT 265  
 DB 309 RAFLSMIDVTGKTKTYTGTGPLEV-YSNWPGEPPNNGAENCVEIFINGQWMDKAGGEQ 367  
 QY 266 MYFVCEP 272  
 DB 368 RLVICEF 374  
 RESULT 4  
 PSPD\_HUMAN  
 ID PSPD\_HUMAN PRT; 375 AA.  
 AC P35247;

DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).  
 GN SFTPD OR SFTPD OR PSPD.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=9315122; PubMed=8428971;  
 RA Crouch E., Rust K., Veile R., Donis-Keller H., Grosso L.;  
 RT "Genomic organization of human surfactant protein D (SP-D). SP-D is  
 RT encoded on chromosome 10q22.2-23.1.";  
 RL J. Biol. Chem. 268:2976-2983(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 214-243.  
 RC TISSUE=Lung, and Amniotic fluid;  
 RX MEDLINE=92322003; PubMed=1339284;  
 RA Lu J., Willis A.C., Reid K.B.M.;  
 RT "Purification, characterization and cDNA cloning of human lung  
 RT surfactant protein D.";  
 RL Biochem. J. 284:795-802(1992).  
 RN [3]  
 RP SEQUENCE OF 60-375 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Lung;  
 RX MEDLINE=91378578; PubMed=1898081;  
 RA Rust K., Grosso L., Zhang V., Chang D., Persson A., Longmore W.,  
 RT "Human surfactant protein D: SP-D contains a C-type lectin  
 RT carbohydrate recognition domain.";  
 RL Arch. Biochem. Biophys. 290:116-126(1991).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=99197291; PubMed=10368295;  
 RA Hakansson K., Lim N.K., Hoppe H.-J., Reid K.B.M.;  
 RT "Crystal structure of the trimeric alpha-helical coiled-coil and the  
 RT three lectin domains of human lung surfactant protein D.";  
 RL Structure 7:255-264(1999).  
 CC -|- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED  
 CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER  
 CC EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE  
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.  
 CC -|- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.  
 CC -|- SUBCELLULAR LOCATION: Extracellular.  
 CC -|- PTM: the N-terminus is blocked.  
 CC -|- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
 CC -|- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: L05485; AAB59450.1;  
 CC EMBL: L05483; AAB59450.1; JOINED.  
 CC EMBL: L05484; AAB59450.1; JOINED.  
 CC EMBL: X65018; CAA46152.1;  
 CC PIR: A45225; A45225.  
 CC PIR: S18382; S18382.  
 CC PDB: 1B08; 29-NOV-99.  
 CC TIM: 178635;  
 CC InterPro: IPR000087; Collagen.  
 CC InterPro: IPR001304; lectin\_c.

[illegible]

Db 134 PGPKEAGPRGVEGAPGMOGSAGAGPAGPKGAPCEQAGPAGNAGAGPAGPQGA 193  
 QY 103 KGLIGIPGKAGTCDGC-----RYKFGQDIDISIAIKTSMKFKVN--VI 149  
 Db 194 PGRGPPGLKGRGAPGRGKIGESGLPSAALRQOMALNGKLORLEAFSRYKKAALF 253  
 QY 150 AGIREHBEFYIVOEKNYRESLTHCRIRGGLAMPKDEAANTLIADYVAKSGFFRVFI 209  
 Db 254 PDGQGVGDKIFRAANSEPFDEKEMCRQAGGOLASPRATENAQQVIVTAHS-KAALF 312  
 QY 210 GVNDLREGEQYFTDPTPLQNNWNEGPPSPDYGHEDCVEMLSGSRWMDTECHLIMYFV 269  
 Db 313 SMIDVTEGKFTPTTGEALV-YSNWAPGPPNNGAENCVEIFITNGQWNRACGEQRLVI 371  
 QY 270 CEF 272  
 Db 372 CEF 374

RESULT 6  
 CONG\_BOVIN  
 ID CONG\_BOVIN STANDARD; PRT; 371 AA.  
 AC P23805;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Conglutinin precursor.  
 GN CGN1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93213261; PubMed=8460993;  
 RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;  
 RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";  
 RL Biochem. Biophys. Res. Commun. 191:335-342(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93277452; PubMed=7684896;  
 RA Lu J., Laursen S.B., Thiel S., Jensenius J.C., Reid B.M.;  
 RT "The cDNA cloning of conglutinin and identification of liver as a  
 primary site of synthesis of conglutinin in members of the Bovidae.";  
 RL Biochem. J. 292:157-162(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94215917; PubMed=8163202;  
 RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,  
 RA Tauber A.I., Sastry K.N.;  
 RT "Bovine conglutinin (BC) mRNA expressed in liver: cloning and  
 characterization of the BC cDNA reveals strong homology to surfactant  
 protein-D.";  
 RL Gene 141:277-281(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94267222; PubMed=8207234;  
 RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,  
 RA Tauber A.I., Sastry K.N.;  
 RT "Bovine conglutinin gene exon structure reveals its evolutionary  
 relationship to surfactant protein-D.";  
 RL J. Immunol. 153:173-180(1994).  
 RN [5]  
 RP SEQUENCE OF 21-371.  
 RX MEDLINE=9131556; PubMed=1993651;  
 RA Lee Y.-M., Leiby K.R., Allar J., Paris K., Lerch B., Okarma T.B.;  
 RT "Primary structure of bovine conglutinin, a member of the C-type  
 animal lectin family.";  
 RL J. Biol. Chem. 266:2715-2723(1991).  
 RN [6]

RP PRELIMINARY SEQUENCE OF 21-52.  
 RX MEDLINE=87184551; PubMed=3566740;  
 RA Young N.M., Leon M.A.;  
 RT "The carbohydrate specificity of conglutinin and its homology to  
 proteins in the hepatic lectin family.";  
 RL Biochem. Biophys. Res. Commun. 143:645-651(1987).  
 CC -!- FUNCTION: CALCIUM-DEPENDENT LECTIN-LIKE PROTEIN WHICH BINDS TO A  
 YEAST CELL WALL EXTRACT & IMMUNE COMPLEXES THROUGH THE COMPLEMENT  
 COMPONENT (C3BI). IT IS CAPABLE OF BINDING NONREDUCING TERMINAL  
 N-ACETYLGLUCOSAMINE, MANNOSE, AND FUCOSE RESIDUES.  
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.  
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D14085; BAA03170.1; -  
 DR EMBL; X71774; CAA50665.1; -  
 DR EMBL; L18871; AAA20126.1; -  
 DR EMBL; U08860; AAB60624.1; -  
 DR EMBL; U08854; AAB60624.1; JOINED.  
 DR EMBL; U08855; AAB60624.1; JOINED.  
 DR EMBL; U08856; AAB60624.1; JOINED.  
 DR EMBL; U08857; AAB60624.1; JOINED.  
 DR EMBL; U08858; AAB60624.1; JOINED.  
 DR EMBL; U08859; AAB60624.1; JOINED.  
 DR PIR; A29416; A29416.  
 DR PIR; A23740; A23740.  
 DR PIR; JN0450; JN0450.  
 DR HSP; P35247; I308.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001304; lectin\_C.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF00059; lectin\_c; 1.  
 DR SMART; SM00034; CLECT; 1.  
 DR PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
 DR PROSITE; PS00041; C-TYPE\_LECTIN\_2; 1.  
 KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;  
 Collagen; Repeat; Calcium; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 371 CONGLUTININ.  
 FT DOMAIN 46 216 COLLAGEN-LIKE.  
 FT DOMAIN 273 371 C-TYPE LECTIN (SHORT FORM).  
 FT MOD\_RES 63 63 HYDROXYLATION.  
 FT MOD\_RES 87 87 HYDROXYLATION.  
 FT MOD\_RES 99 99 HYDROXYLATION.  
 FT MOD\_RES 135 135 HYDROXYLATION.  
 FT MOD\_RES 141 141 HYDROXYLATION.  
 FT MOD\_RES 159 159 HYDROXYLATION.  
 FT MOD\_RES 162 162 HYDROXYLATION.  
 FT MOD\_RES 198 198 HYDROXYLATION.  
 FT SITE 201 203 CELL ATTACHMENT SITE (POTENTIAL).  
 FT DISULFID 275 369 BY SIMILARITY.  
 FT DISULFID 347 361 BY SIMILARITY.  
 FT CARBOHYD 337 337 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CONFLICT 173 173 R -> H (IN REF. 2 AND 3).  
 FT CONFLICT 210 210 K -> S (IN REF. 5).  
 FT CONFLICT 218 218 V -> A (IN REF. 2).  
 FT CONFLICT 272 272 E -> V (IN REF. 2).  
 SQ SEQUENCE 371 AA; 37994 MW; 867BB41992544B1F CRC64;

Query Match 22.1%; Score 328.5; DB 1; Length 371;  
 Best Local Similarity 29.5%; Pred. No. 8e-21;  
 Matches 85; Conservative 35; Mismatches 95; Indels 73; Gaps 10;

QY 46 PGPKGDDGEKGDGP-----CEEKGKGVGRMPKGIKGEIGMDRGNIGK-----TGP 93

Db	96	PPKGDGTGPKGPEWPGPAGREGPSKQSGMSPETGPKGKTGPKGVGAPGQGFQCP	153
QY	94	ICKKDGKRGKLLIGPG-	116
Db	156	SLKKGKGAQGTGAPGRAGVTPSGAIGPOGPGSANGPPGLKGRDGPCTGAKGESGL	215
QY	117	-----TWCDGRRYKRVFQGLDSTANKTSKMKFKVKNVIAGIRTEKEFYIVQEE	166
Db	216	AEVNALKQRVTILD-GHLRFQN-----AFSYKKAVLFFDGOAVG-----EKIFTAGAV	265
QY	167	KNYRESLTHCRINGGMLAMPKDEAANTLIADYVAKSGFERFVGWDLEREGQYFTDNT	226
Db	266	KYSYDAQLCRKAGQLASPRSSAENEAVTQW-RAQEKNAVLSMNDISTSGRFTYPTGE	324
QY	227	PLQNYSNWEGEP--SDPYGHEDCEVMLSSGRWNNDTECHLTMYFYCFE	272
Db	325	ILV-YSNWADGPPNNSDEGQENCVEIFPDGKWNVPCKSQLLYICEF	371
RESULT	7		
PSPA	RABIT		
ID	PSPA_RABIT	STANDARD;	PRT; 247 AA.
AC	P12842;		
DT	01-OCT-1989	(Rel. 12, Created)	
DT	01-OCT-1989	(Rel. 12, Last sequence update)	
DT	16-OCT-2001	(Rel. 40, Last annotation update)	
DE	Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A) (PSPN).		
DE	SF1PAL OR SF1PA OR SF1P1.		
OS	Oryctolagus cuniculus (Rabbit).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		
ON	NCBI_TaxID=9986;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=88139348; PubMed=2830270;		
RA	Boggarum V., Qing K., Mendelson C.R.;		
RA	"The major apoprotein of rabbit pulmonary surfactant. Elucidation of		
RT	primary sequence and cyclic AMP and developmental regulation.";		
RL	J. Biol. Chem. 263:2939-2947(1988).		
RL	[2]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN=NEW ZEALAND WHITE; TISSUE=Liver;		
RX	MEDLINE=92312742; PubMed=1616051;		
RA	Chen Q., Boggarum V., Mendelson C.R.;		
RT	"Rabbit lung surfactant protein A gene: identification of a lung-		
RL	specific DNase I hypersensitive site.";		
RL	Am. J. Physiol. 262:L662-L671(1992).		
CC	-!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSPA BINDS TO SURFACTANT		
CC	PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACTANT TENSION AT THE		
CC	AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS		
CC	ESSENTIAL FOR NORMAL RESPIRATION.		
CC	-!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.		
CC	-!- SUBCELLULAR LOCATION: Extracellular.		
CC	-!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%		
CC	PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,		
CC	CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL		
CC	HYDROPHOBIC PROTEINS (SP-B AND SP-C).		
CC	-!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.		
CC	-!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> )		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .		
DR	EMBL; J03542; AAA31465.1; -;		
DR	EMBL; L19387; AAA31468.1; -;		
DR	PIR; A29931; LNRBPS.		



RX MEDLINE=87137502; PubMed=3029088;  
RA Drickamer K., McCreary V.;  
RT "Exon structure of a mannose-binding protein gene reflects its  
RT evolutionary relationship to the asialoglycoprotein receptor and  
RT nonfibrillar collagens";  
RL J. Biol. Chem. 262:2582-2589(1987).  
RN [3]  
RP SEQUENCE OF 18-42.  
RX MEDLINE=87222358; PubMed=3584121;  
RA Ikeda K., Sannoh T., Kawasaki N., Kawasaki T., Yamashina I.;  
RT "Serum lectin with known structure activates complement through the  
RT classical pathway";  
RL J. Biol. Chem. 262:7451-7454(1987).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 124-238.  
RX MEDLINE=92086855; PubMed=1721241;  
RA Weis W.I., Kahn R., Fourme R., Drickamer K., Hendrickson W.A.;  
RT "Structure of the calcium-dependent lectin domain from a rat mannose-  
RT binding protein determined by MAD phasing.";  
RL Science 254:1608-1615(1991).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 124-238.  
RX MEDLINE=93063338; PubMed=1436090;  
RA Weis W.I., Drickamer K., Hendrickson W.A.;  
RT "Structure of a C-type mannose-binding protein complexed with an  
RT oligosaccharide.";  
RL Nature 360:127-134(1992).  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 84-238.  
RX MEDLINE=95219384; PubMed=7704532;  
RA Weis W.I., Drickamer K.;  
RT "Trimeric structure of a C-type mannose-binding protein.";  
RL Structure 21:1227-1240(1994).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 90-238.  
RX MEDLINE=99119227; PubMed=9922165;  
RA Ng K.K.-S., Park-Snyder S., Weis W.I.;  
RT "Ca2+-dependent structural changes in C-type mannose-binding  
RT proteins.";  
RL Biochemistry 37:17965-17976(1998).  
CC -!- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-  
CC DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,  
CC BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF  
CC THE ANTIBODY.  
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.  
CC -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST  
CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI  
CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND  
CC LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNAL SPACE  
CC OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE PROTEIN.  
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: M14105; AAA98781.1;  
CC EMBL: M14104; AAA98781.1; JOINED.  
CC FIR: B24791; LNR7MA.  
CC DR PDB: 1NSB; 15-JAN-92.  
CC DR PDB: 2MSB; 31-OCT-93.  
CC DR PDB: 1AFA; 03-APR-96.  
CC DR PDB: 1AFB; 03-APR-96.  
CC DR PDB: 1AFD; 03-APR-96.  
CC DR PDB: 1FTW; 07-FEB-95.  
CC DR PDB: 1KMB; 12-FEB-97.  
CC DR PDB: 2KMB; 12-FEB-97.  
CC DR PDB: 3KMB; 12-FEB-97.

DR PDB: 4KMB; 12-FEB-97.  
DR PDB: 1YTT; 10-JUN-96.  
DR PDB: 1BCH; 17-JUN-98.  
DR PDB: 1BCU; 17-JUN-98.  
DR PDB: 1BUU; 09-SEP-98.  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR001304; lectin\_c.  
DR Pfam: PR00059; lectin\_c; 1.  
DR SMART: SM00034; CLECT; 1.  
DR PROSITE: PS00615; C-TYPE\_LLECTIN\_1; 1.  
DR PROSITE: PS00041; C-TYPE\_LLECTIN\_2; 1.  
DR Lectin; Hydroxylation; Liver; Glycoprotein; Mannose-binding; Membrane;  
KW Calcium; Collagen; Repeat; Signal; 3D-structure.  
FT SIGNAL 1 17  
FT CHAIN 18 238  
FT DOMAIN 39 88  
FT MOD\_RES 43 43  
FT MOD\_RES 61 61  
FT MOD\_RES 67 67  
FT MOD\_RES 73 73  
FT MOD\_RES 78 78  
FT DISULFID 145 234  
FT DISULFID 212 226  
FT CONFLICT 156 156  
FT STRAND 127 136  
FT HELIX 138 147  
FT TURN 148 149  
FT STRAND 151 152  
FT HELIX 158 168  
FT STRAND 172 177  
FT TURN 182 183  
FT STRAND 186 187  
FT TURN 188 189  
FT STRAND 192 192  
FT STRAND 198 198  
FT TURN 200 201  
FT TURN 207 208  
FT STRAND 212 215  
FT HELIX 217 219  
FT STRAND 221 224  
FT TURN 226 227  
FT STRAND 230 237  
SQ SEQUENCE 238 AA; 25308 MW; 1A927482B9A8CB3D CRC64;  
  
Query Match 19.5%; Score 290; DB 1; Length 238;  
Best Local Similarity 28.7%; Pred. No. 9.3e-18;  
Matches 75; Conservative 49; Mismatches 105; Indels 32; Gaps 8;  
  
QY 14 ILVLFLQLQISLGLDIDSDRPTAEVCATHTISPGKDDGKDPGEGKHGKVGKMGPK 73  
DB 6 LLVLLCVSVSSSSQ--TCEETLKTC--VIAGRGRGCPKGEKGPQG-----GLR 55  
QY 74 GKIGELGDMGRGNIGTKTGPICKGDKGKGLLGPCKEKGAGTVCCGGRYKFGVGLDI 133  
DB 56 GLQGPPLKGLPPGSGVAPGSGQPGKQKGRG-----DSRAIEVKLANMEA 100  
QY 134 STARKTSMKFKVKNVITAGI--RETEKEFYIVOEKNYRESLTHCRTRGGMAMPKDEAA 191  
DB 101 EINTLRSKLEITNKLHAFSGKSGKGFVTNHPFFSKVKALCSELRTGTVAIIPNAEE 160  
QY 192 NTILIADYVAKSGFRVFIGVNDLEREGQYMTDNTPLQYNSWNEGEPSPYGHEDCYEM 251  
DB 161 NKAIOE-VAKTS---AFLGITDEVTEGQFMVYVGGRL-TYSNKKKDEPDHSGEDCVTI 215  
QY 252 LSSGRWNDETECHLTMYVFCVF 272  
DB 216 VDNLGWNDISQASHTAVCEVF 236

RESULT 9  
MABC\_MOUSE

ID MABC\_MOUSE STANDARD; PRT; 244 AA.  
AC P41317;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)  
DE (RA-reactive factor P28A subunit) (RARE/P28A).  
GN MBL2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CRA/J; TISSUE=Liver;  
RA MEDLINE=91302823; PubMed=1712818;  
RA Sastri K., Zahedi K., Lelias J.M., Whitehead A.S., Ezekowitz R.A.;  
RT "Molecular characterization of the mouse mannose-binding proteins."  
RT The mannose-binding protein A but not C is an acute phase reactant.";  
RL J. Immunol. 147:692-697(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RA MEDLINE=95284466; PubMed=7766991;  
RA Sastri R., Wang J.S., Brown D.C., Ezekowitz R.A., Tauber A.I.,  
RA Sastri K.N.;  
RT "Characterization of murine mannose-binding protein genes Mbl1 and  
RT Mbl2 reveals features common to other collectin genes.";  
RL Mamm. Genome 6:103-110(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kuge S., Ihara S., Watanabe E., Watanabe M., Takishima K., Suga T.,  
RA Manaiya G., Kawakami M.;  
RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-  
CC DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,  
CC BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF  
CC THE ANTIBODY.  
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.  
CC -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST  
CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI  
CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND  
CC LYSOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNAL  
CC SPACE OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE  
CC PROTEIN  
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; S42294; AAB19343.1;  
CC EMBL; U09016; AAB20210.1;  
CC EMBL; U09013; AAB20210.1; JOINED.  
CC EMBL; U09014; AAB20210.1; JOINED.  
CC EMBL; U09015; AAB20210.1; JOINED.  
CC EMBL; D11440; BAA02005.1;  
CC HSSP; P08661; 1R00.  
CC MGD; MGI:96924; Mbl2.  
CC InterPro: IPR000087; Collagen.  
CC InterPro: IPR001304; lectin\_c.  
CC Pfam: PF00059; lectin\_c; 1.  
CC SMART; SM00034; CLECT; 1.  
CC PROSITE; PS00615; C-TYPE LECTIN.1; 1.  
CC PROSITE; PS00615; C-TYPE LECTIN.2; 1.  
CC LECTIN; Hydroxylation; Liver; Glycoprotein; Mannose-binding; Membrane;  
CC Calcium; Collagen; Repeat; Signal.  
CC SIGNAL 1 18 BY SIMILARITY.

FT CHAIN 19 244  
FT DOMAIN 38 96  
FT C-TYPE LECTIN (SHORT FORM).  
FT MOD\_RES 43 43  
FT MOD\_RES 58 58  
FT MOD\_RES 69 69  
FT MOD\_RES 78 78  
FT MOD\_RES 81 81  
FT DISULFID 29 29  
FT DISULFID 34 34  
FT DISULFID 151 240  
FT DISULFID 218 232  
FT CONFLICT 3 3  
FT CONFLICT 15 15  
SQ SEQUENCE 244 AA; 25957 MW; 49AE84E2290DEB0A CRC64;  
I -> L (IN REF. 1).  
V -> A (IN REF. 1).

Query Match 19.5%; Score 290; DB 1; Length 244;  
Best Local Similarity 28.8%; Pred. No. 9.5e-18;  
Matches 80; Conservative 56; Mismatches 100; Indels 42; Gaps 11;

QY 1 MNCFASILLRNQFILLVLFLLQISGLDIDSRPAEVCATHTIS--PGPKGDDGKGP 58  
Db 1 MSIFTSFL-----LLCVTVVTAETLTGQVQNSCPVYTCSFGLNGEPFGKDRGAKGEK 55  
QY 59 GREGKHGKVGKMGPKIGKELGMDGRNIGTKPTGPKGKDGKGLLGPGEKAGTV 118  
Db 56 GEPGQ---GLRGLOGPKGKVGTPGPNGLGKAGVGPKGDRGDR----- 97  
QY 119 CDGGRKRF-VGQIDISIAKLTKSMKFKVN-VIAGIRETEEFKYYIVQEEKYRESL-TH 175  
Db 98 -----EPTSEIDSEIAALSELRAIRNWLFLSELSEKVGKTYFVSVKMSLDVRKAL 150  
QY 176 CRIRGGLAMPKDEAANTLIADYVAKSGFRRVIGVNDLEREGYMTDNTPLQ-NYSNW 234  
Db 151 CSEFGSVAPRPAENSAI-OKVAKD---IAYLGITDVRVEGS--FEDLTGRRVRYTNW 204  
QY 235 NEGEPSDPYGHEDCVEMLSGSGRWNTCHLTMYFVCEP 272  
Db 205 NDEGPNNTGDEGCVVILGNKGKNDVPCSDSFLAICEF 242

RESULT 10  
MABC\_HUMAN STANDARD; PRT; 248 AA.  
ID MABC\_HUMAN  
AC P11226;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mannose-binding protein C precursor (MBP-C) (MBP1) (Mannan-binding  
DE protein) (Mannose-binding lectin).  
GN MBL2 OR MBL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=90010778; PubMed=2477486;  
RA Sastri K., Herman G.A., Day L.E., Deignan E., Bruns G., Morton C.C.,  
RA Ezekowitz R.A.B.;  
RT "The human mannose-binding protein gene. Exon structure reveals its  
RT evolutionary relationship to a human pulmonary surfactant gene and  
RT localization to chromosome 10.";  
RL J. Exp. Med. 170:1175-1189(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86171281; PubMed=2450948;  
RA Ezekowitz R.A.B., Day L.E., Herman G.A.;  
RT "A human mannose-binding protein is an acute-phase reactant that  
RT shares sequence homology with other vertebrate lectins.";  
RL J. Exp. Med. 167:1034-1046(1988).

[3] SEQUENCE FROM N.A.  
RX MEDLINE=90073571; PubMed=2590164;  
RA Taylor M.E., Brickell P.M., Craig R.K., Summerfield J.A.;  
RA "Structure and evolutionary origin of the gene encoding a human serum  
RT mannose-binding protein.";  
RL Biochem. J. 262:763-771(1989).  
[4]  
RN PARTIAL SEQUENCE.  
RN TISSUE=Liver, and Plasma:  
RC MEDLINE=95073978; PubMed=7982896;  
RX Kurata H., Sannoh T., Kozutsumi Y., Yokota Y., Kawasaki T.;  
RA "Structure and function of mannan-binding proteins isolated from  
RT human liver and serum.";  
RL J. Biochem. 115:1148-1154(1994).  
[5]  
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 108-248.  
RN MEDLINE=95360730; PubMed=7634089;  
RX Sheriff S., Chang C.Y., Ezekowitz R.A.;  
RA "Human mannose-binding protein carbohydrate recognition domain  
RT trimerizes through a triple alpha-helical coiled-coil.";  
RL Nat. Struct. Biol. 1:783-794(1994).  
[6]  
RN VARIANT ASP-54.  
RN MEDLINE=91269930; PubMed=1675710;  
RX Sumiya M., Super M., Tabona P., Levinsky R.J., Arai T., Turner M.W.,  
RA Summerfield J.A.;  
RL "Molecular basis of opsonic defect in immunodeficient children.";  
RN Lancet 337:1569-1570(1991).  
[7]  
RN VARIANTS ASP-54 AND GLU-57.  
RN MEDLINE=93258313; PubMed=1304173;  
RX Lipscombe R.J., Sumiya M., Hill A.V.S., Lau Y.L., Levinsky R.J.,  
RA Summerfield J.A., Turner M.W.;  
RL "High frequencies in African and non-African populations of  
RT independent mutations in the mannose binding protein gene.";  
RN Hum. Mol. Genet. 1:709-715(1992).  
[8]  
RN ERRATUM.  
RX Lipscombe R.J., Sumiya M., Hill A.V.S., Lau Y.L., Levinsky R.J.,  
RA Summerfield J.A., Turner M.W.;  
RL Hum. Mol. Genet. 2:342-342(1993).  
[9]  
RN VARIANT ASP-54.  
RX MEDLINE=93265124; PubMed=1303250;  
RA Super M., Gillies S.D., Foley S., Sastry K., Schweinle J.E.,  
RA Silverman V.J., Ezekowitz R.A.;  
RL "Distinct and overlapping functions of allelic forms of human mannose  
RT binding protein.";  
RN Nat. Genet. 2:50-55(1992).  
[10]  
RN VARIANTS CYS-52; ASP-54 AND GLU-57.  
RX MEDLINE=93374928; PubMed=10447262;  
RA Gbolade M., Muralitharan S., Besmond C.;  
RL "Genotyping of the three major allelic variants of the human  
RT mannose-binding lectin gene by denaturing gradient gel  
RT electrophoresis.";  
RN Hum. Mutat. 14:80-83(1999).  
CC -1- FUNCTION: BINDS MANNOSE AND N-ACETYLGALUCOSAMINE IN A CALCIUM-  
CC DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,  
CC BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF  
CC THE ANTIBODY  
CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.  
CC -1- DISEASE: THERE IS AN ASSOCIATION BETWEEN LOW LEVELS OF MBL AND A  
CC DEFECT OF OPSONIZATION WHICH RESULTS IN SUSCEPTIBILITY TO FREQUENT  
CC AND CHRONIC INFECTIONS.  
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN  
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -1- CAUTION: THERE ARE EXTENSIVE DIFFERENCES BETWEEN THE REVISED  
CC SEQUENCE IN REF.1 AND THAT PUBLISHED IN REF.2.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

## RESULT 11

PSPA\_PIG STANDARD: PRT; 249 AA.  
AC P49874;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A) (PSPAP).  
GN SFTPA1 OR SFTPA OR SFTPL.  
OS Sus scrofa (Pig).  
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RA Adamou J.E., Elshourbagy N.A.; Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
CC !- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSPAP BINDS TO SURFACTANT PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS ESSENTIAL FOR NORMAL RESPIRATION.  
CC !- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.  
CC !- SUBCELLULAR LOCATION: Extracellular.  
CC !- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL HYDROPHOBIC PROTEINS (SP-B AND SP-C) (BY SIMILARITY).  
CC !- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
CC !- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; LA1350; AAA88403.1; .  
CC HSP; P19999; 1YTT.  
CC InterPro: IPR000087; Collagen.  
CC InterPro: IPR001304; lectin\_c.  
CC Pfam: PF00059; lectin\_c; 1.  
CC SMART; SM00034; CLECT; 1.  
CC PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
CC PROSITE; PS50041; C-TYPE\_LECTIN\_2; 1.  
CC Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation; Signal; Lectin; Collagen; Repeat.  
CC SIGNAL 1 20 POTENTIAL.  
CC CHAIN 21 249 PULMONARY SURFACTANT-ASSOCIATED PROTEIN A.  
CC DOMAIN 28 100 COLLAGEN-LIKE.  
CC FT DOMAIN 153 249 C-TYPE LECTIN (SHORT FORM).  
CC FT DISULFID 155 247 BY SIMILARITY.  
CC FT DISULFID 225 239 BY SIMILARITY.  
CC FT CARBOHYD 208 208 N-LINKED (GLCNAC...) (POTENTIAL).  
CC SEQUENCE 249 AA; 26702 MW; 3C4E05AD07F2A7CD CRG64;

## Query Match

Best Local Similarity 19.28; Score 285.5; DB 1; Length 249;  
Matches 85; Conservative 35; Mismatches 11; Indels 41; Gaps 12;

QY 14 ILLVFLQLQSLGLDIDSRPTAEVCAVTHHTISPGKGDGKPGEGKGKVGMRGPK 73  
Db 6 LALFLLVSLGLECDV-----KEVCLA---SFGIPGTPSHGLPGRDGRDGIKDFGPP 57  
QY 74 GKGLGDMGDRGNIGTKGPIGKKGKGLGIPGEGKAGKASTCDCGRYKRFVGLDI 133  
Zb 58 GPMGPPGMA--GPPQDGMICAPGLPGEKGEKGPGRGPPG-----LPAHLD- 104

QY 134 SIARLKTSKFKVKNVI---AGIRETEEFYIVQEKNYRE-----SLHHC-RING 180  
Db 105 --EELQSALHEIRHQILQSMGLVLFQE--FMLAVGEKVFSTNGQSVAFWWSLSCVPEQV 160  
QY 181 GMLAMPKDEARANTLIADYAKSGFFRFGVIGVNDLREGQYMETDNTPLQYNSWNGEPPS 240  
Db 161 GRIAAPRSFEENAIASIVKKHNTY-AYLGLVSPTAGDFFYLDGTPV-NYTNWYGPGR 218  
QY 241 DPYGHEDCVESLSSGRWNDECHLTMYEVCDF 272  
Db 219 G-RGKEKCVEMTYDQWDRNCQQLALICEF 249

## RESULT 12

MABC\_BOVIN STANDARD; PRT; 249 AA.  
AC 002659;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein).  
GN MBL.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=97228413; PubMed=9074491;  
RA Kawai T., Suzuki Y., Eda S., Ohtani K., Kase T., Fujinaga Y., Sakamoto T., Kurimura T., Wakamiya N.;  
RT "Cloning and characterization of a cDNA encoding bovine mannan-binding protein."  
RL Gene 186:161-165(1997).  
CC !- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS, BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF THE ANTIBODY (BY SIMILARITY).  
CC !- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.  
CC !- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
CC !- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D73408; BAA18935.1; .  
CC HSP; P11226; 1HUP.  
CC InterPro: IPR000087; Collagen.  
CC InterPro: IPR001304; lectin\_c.  
CC Pfam: PF00059; lectin\_c; 1.  
CC SMART; SM00034; CLECT; 1.  
CC PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
CC PROSITE; PS50041; C-TYPE\_LECTIN\_2; 1.  
CC Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane; Calcium; Collagen; Repeat; Signal.  
CC SIGNAL 1 20 POTENTIAL.  
CC CHAIN 21 249 MANNOSE-BINDING PROTEIN C.  
CC DOMAIN 43 101 COLLAGEN-LIKE.  
CC FT DOMAIN 154 247 C-TYPE LECTIN (SHORT FORM).  
CC FT MOD\_RES 48 48 HYDROXYLATION (POTENTIAL).  
CC FT MOD\_RES 63 63 HYDROXYLATION (POTENTIAL).  
CC FT MOD\_RES 74 74 HYDROXYLATION (POTENTIAL).  
CC FT MOD\_RES 83 83 HYDROXYLATION (POTENTIAL).  
CC FT MOD\_RES 86 86 HYDROXYLATION (POTENTIAL).  
CC FT DISULFID 34 34 INTERCHAIN (BY SIMILARITY).



Db 60 MGPPGEMPCPGNDGLPGAPGPGCEGK---GPGGPGGLPAHLDE-----ELOAT 110  
QY 135 IARLKTSMKVKVNTA---GIRETEKFFYIVQEEKYRSLRTHCRGGMAMPKDEAA 191  
Db 111 LHDFRHQILQTRGALSLOGSMTVGVKFSNGSQSITFDIAQACARAGGRIAPRNP 170  
QY 192 NTLIADYVAKSGFFRVFVGVNDLREGQVMTDNTPLQNSNNNEGSDPYGHEDQVEM 251  
Db 171 NEATASFVKYNTY-AYVGLTEGSPGDFRYSDDGTPV-NVTNWRGEFAG-RGKEQCQVEM 227  
QY 252 LSSGRWNTDCHLTWYFVCEP 272  
Db 228 YTDGQWNRNCLYSRLICEF 248

RESULT 14  
PSPA\_CANFA  
ID PSPA\_CANFA STANDARD; PRT; 248 AA.  
AC P06908;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 46, Last annotation update)  
DE Pulmonary surfactant-associated protein A precursor (SP-A) (PSP-A)  
DE (PSP-A).  
OS SFTPAL OR SFTPA OR SFTPL.  
GN Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A., AND SIGNAL SEQUENCE CLEAVAGE SITE.  
RX MEDLINE=86016705; PubMed=3863100;  
RA Benson B., Hawgood S., Schilling J., Clements J., Damm D., Cordell B.,  
RA White R.T.;  
RT "Structure of canine pulmonary surfactant apoprotein: cDNA and  
RT complete amino acid sequence."  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6379-6383(1985).  
RN [2]  
RP C-TYPE LECTIN DOMAIN.  
RX MEDLINE=87115634; PubMed=3808053;  
RA Parthy L.;  
RT "Is lung surfactant protein a lectin-collagen hybrid?";  
RL Nature 325:490-490(1987).  
CC -!- FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT  
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE  
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS  
CC ESSENTIAL FOR NORMAL RESPIRATION.  
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; M11769; AAA30887.1;  
CC PIR; A25296; LNDGSP.  
CC HSSP; P35247; 1B08.  
CC InterPro; IPR000087; Collagen.  
CC InterPro; IPR001304; lectin\_C.  
CC Pfam; PF01391; Collagen; 1.  
CC Pfam; PF00059; lectin\_C; 1.  
CC SMART; SM00034; CLECT; 1.  
CC PROSITE; PS00615; C-TYPE LECTIN\_1; 1.

DR PROSITE; PS00441; C-TYPE LECTIN\_2; 1.  
KW Glycoprotein; Calcium; Surface film; Caseous exchange; Hydroxylation;  
KW Signal; Lectin; Collagen; Repeat.  
FT SIGNAL 1 17  
FT CHAIN 18 248 PULMONARY SURFACTANT-ASSOCIATED PROTEIN  
FT A.  
FT DOMAIN 28 100 COLLAGEN-LIKE.  
FT DOMAIN 153 248 C-TYPE LECTIN (SHORT FORM).  
FT DISULFID 155 246 BY SIMILARITY.  
FT DISULFID 224 238 BY SIMILARITY.  
FT CARBOHYD 20 20 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 207 207 N-LINKED (GLCNAC... ) (PROBABLE).  
SQ SEQUENCE 248 AA; 26268 MW; 340FE95D4E2502C0 CRC64;

Query Match 18.9%; Score 280; DB 1; Length 248;  
Best Local Similarity 30.0%; Pred. No. 6.9e-17;  
Matches 81; Conservative 36; Mismatches 115; Indels 38; Gaps 11;  
QY 14 ILLVFLIQSGLDIDSRPTAEVCATHITISPGKGGDGEKGPGEKHKGVGMGPK 73  
Db 6 LALATLLMVSGI-----ENNTKDVCGV---NPGIPGTPGSHGLPGRDGRGVKGP 57  
QY 74 GIKGELGDM-GDRGNIGKTGPIGKKGDKGKGLLGPGEKAGTVCDCGRYKFKVG 132  
Db 58 GPLGPGGMPGHPGPMGTGAPGVAGERGEK---GPGGPGP-----LPASLD 104  
QY 133 ISIALKTSMKFVKVNI---AGIRETEE-----KFYIVQEEKYRSLRTHCRIGGM 182  
Db 105 ---EELQTLHLRHQILQTMGVLSHESLLVGVKVFSSNAQSFNDIQELCAGAGG 161  
QY 183 LAMPKDEANTLIADYVAKSGFFRVFVGVNDLREGQVMTDNTPLQNSNNNEGSDP 242  
Db 162 IAAPMSPEENEAIVKVKYNTY-AYLGLVESPDGQFYMDGAPV-NVTNYPGEPRG- 218  
QY 243 YGHEDQVEMLSGRWNTDCHLTWYFVCEP 272  
Db 219 RGKEQCQVEMTQGWNNKCLQYRLICEF 248

RESULT 15  
MABC\_RAT  
ID MABC\_RAT STANDARD; PRT; 244 AA.  
AC P08661;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Mannose-binding protein C precursor (MBP-C) (Mannan-binding protein)  
DE (RA-reactive factor P28A subunit) (RAFP/P28A).  
GN MBL2  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=101116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=92299655; PubMed=1607365;  
RA Wada M., Itoh N., Ohta M., Kawasaki T.;  
RT "Characterization of rat liver mannan-binding protein gene."  
RL J. Biochem. 111:66-73(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=86196130; PubMed=3009480;  
RA Drickamer K., Dordal M.S., Reynolds L.;  
RT "Mannose-binding proteins isolated from rat liver contain  
RT carbohydrate-recognition domains linked to collagenous tails.  
RT Complete primary structures and homology with pulmonary surfactant  
RT apoprotein."  
RL J. Biol. Chem. 261:6878-6887(1986).  
RN [3]  
RP SEQUENCE FROM N.A.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:31:36 ; Search time 20.65 Seconds  
(without alignments)  
327.646 Million cell updates/sec

Title: US-09-600-932-2  
Perfect score: 1484  
Sequence: 1 MNGFASLLRNQFILLVLF.....NDTECHLTMVFCFIRKKK 277

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	328.5	22.1	351	3	US-09-011-735-1
2	289.5	19.5	248	4	US-09-198-603C-2
3	191.5	12.9	1024	3	US-08-931-820-1
4	191.5	12.9	1366	3	US-08-963-825-19
5	191.5	12.9	1366	4	US-09-570-573-19
6	191.5	12.9	1366	4	US-09-548-608-19
7	187.5	12.6	489	2	US-08-794-795-7
8	187.5	12.6	489	4	US-09-249-200-7
9	187.5	12.6	518	1	US-08-393-367B-2
10	187.5	12.6	518	3	US-08-893-467A-2
11	187	12.6	161	3	US-09-011-735-6
12	181.5	12.2	128	4	US-09-227-357-190
13	180	12.1	107	6	5514582-17
14	180	12.1	557	3	US-09-320-095-10
15	180	12.1	557	4	US-09-523-487-10
16	179	12.1	495	2	US-08-794-795-2
17	179	12.1	495	4	US-09-249-200-2
18	179	12.1	520	2	US-08-794-795-6
19	179	12.1	520	4	US-09-249-200-6
20	177.5	12.0	1060	3	US-08-931-820-3
21	177.5	12.0	1418	3	US-08-963-825-20
22	177.5	12.0	1418	4	US-09-010-959-1
23	177.5	12.0	1418	4	US-09-570-573-20
24	177.5	12.0	1418	4	US-09-548-608-20
25	177	11.9	294	4	US-09-188-930-294
26	176.5	11.9	532	1	US-08-494-168-9
27	176.5	11.9	595	4	US-09-219-849-48

28	176.5	11.9	595	4	US-09-219-849-50
29	176.5	11.9	822	4	US-09-219-849-49
30	175.5	11.8	1057	3	US-08-931-820-1
31	175.5	11.8	1341	3	US-08-963-825-18
32	175.5	11.8	1341	4	US-09-570-573-18
33	175.5	11.8	1341	4	US-09-548-608-18
34	175	11.8	171	3	US-09-011-735-2
35	175	11.8	1442	2	US-08-316-650-12
36	175	11.8	1442	5	PCT-US95-02251-12
37	174.5	11.8	125	3	US-08-722-126A-7
38	174.5	11.8	125	3	PCT-US95-04258-7
39	174.5	11.8	287	1	US-08-365-103B-4
40	174.5	11.8	300	1	US-08-365-103B-6
41	174.5	11.8	327	1	US-08-365-103B-2
42	174	11.7	1057	3	US-08-931-820-4
43	174	11.7	1078	3	US-08-963-825-21
44	174	11.7	1078	4	US-09-570-573-21
45	174	11.7	1078	4	US-09-548-608-21

ALIGNMENTS

RESULT 1  
US-09-011-735-1  
; Sequence 1, Application US/09011735B  
; Patent No. 6110708  
; GENERAL INFORMATION:  
; APPLICANT: Wakamiya, No. 6110708Utaka  
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof  
; FILE REFERENCE: 19036/34548  
; CURRENT APPLICATION NUMBER: US/09/011.735B  
; EARLIER FILING DATE: 1998-05-22  
; EARLIER APPLICATION NUMBER: JP 7-209698  
; EARLIER FILING DATE: 1995-08-17  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Bovine  
US-09-011-735-1

Query Match	22.1%	Score 328.5	DB 3	Length 351
Best Local Similarity	29.2%	Pred. No. 4.1e-25		
Matches	84	Conservative	33	Mismatches 96; Indels 73; Gaps 9
Qy	46	PGPKGDDGEGKDP-----GEGKHGKVGKRGPKGKIGELGMDGRGNICK-----TGP	93	
Db	76	PGPKGDTGPRGPGMPGPGAGREGPSGKQSGMPPGTPGPKGTGPKGGVAGPGIQGFGP	135	
Qy	94	IGKKGDKGEKGLGIPGEGKAG-----	116	
Db	136	SLKGEKAGETGAGRAGVTGPGSAGIGGPGSGARGPPGLKGRDGDPTGAGSGSL	195	
Qy	117	-----TVCDCGRYKFKGQIDISIAIKTSMKFKVKNVIAGIRETEEKFYIVQEE	166	
Db	196	AEVNALKQRVILD-GHLRRFQN---AFSQYKAVLPDQAVG-----EKIFKTAGV	245	
Qy	167	KNYRSLTHCRIRGGMALPKDEAANLIADYVAKSGFFRFVIGVNLDEREGQYMTNT	226	
Db	246	KSYSDAEOLCREAGOLASPRSAENAVTMV-RAQEKNAVLSMNDISTEGRFTYPTGE	304	
Qy	227	PLQYNSNNKEGP--SDPYGHEDCVEMLSGRWNDTECHLTMVFCFEE	272	
Db	305	ILV-YSNWADGEPNNSDEQPCNVEIFPDGKWNVDVPCSKOLLVICEF	351	

RESULT 2  
US-09-198-603C-2  
; Sequence 2, Application US/09198603C  
; Patent No. 6337193



```

; GENERAL INFORMATION:
; APPLICANT: TULLY, Raymond E.
; APPLICANT: CALTAGIRONE, G. Thomas
; APPLICANT: MOYER, Shawn S.
; APPLICANT: RONNING, Michael T.
; TITLE OF INVENTION: EXPRESSION OF MANNOSE-BINDING PROTEIN IN METHYLOTROPHIC
; TITLE OF INVENTION: YEAST
; FILE REFERENCE: A7290
; CURRENT APPLICATION NUMBER: US/09/198,603C
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Human
; US-09-198-603C-2

```

[illegible]

```

RESULT 3
US-08-931-820-2
: Sequence 2, Application US/08931820
: Patent No. 6010863
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Assay for collagen degradation
: NUMBER OF SEQUENCES: 4
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/931,820
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP 96202596.1
: FILING DATE:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1024 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:

```

```
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-2

Query Match      12.9%; Score 191.5; DB 3; Length 1024;
Best Local Similarity 43.5%; Pred.No. 1.1e-10;
Matches 40; Conservative 9; Mismatches 22; Indels 21; Gaps 2;
```

QY 46 PGPAGDDGEGKDPGEGRGHKGKVGWRMPFGIKGLSDMDRGNICKGTPTKK----- 97  
||||| : ||||| : ||||| : ||||| :

Dd 479 PGPSGAGEVCKTGERLGHGFPGPAGPRGRPPGESGAAGTPTIGSRGPSGPFG 538

QY 98 -CDKGEKGLI-----GIPGEKGVAG 116  
:||||| : ||||| : ||||| :

Dd 539 DGNKGEPGVAVGTAGPSGGSPGLFGRGAAG 570

RESULT 4  
US-08-963-825-19  
; Sequence 19, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Ovist, Per  
; APPLICANT: Bonder, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,825  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187,319  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: collagen alpha 2- type I  
; US-08-963-825-19

Query Match 12.9%; Score 191.5; DB 3; Length 1366;  
Best Local Similarity 43.5%; Pred. No. 1.6e-10;  
Matches 40; Conservative 9; Mismatches 22; Indels 21; Gaps 2;

```

QY      98  -GDGKEKGL-----GIPCEKKGAG 116
         |...|...|
         |...|...|
Db      618  DGNKGEPGVGVGAVGTAGPSGSLFEERGAAG 649

RESULT  7
US-08-794-795-7
; Sequence 7, Application US/08794795
; Patent No. 5916766
; GENERAL INFORMATION:
; APPLICANT: Elshourlagy, Nabil

```

```

? ZIP: 19482
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? FILING DATE: 12-FEB-1999
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/794,795
? FILING DATE: 04-FEB-1997
? APPLICATION NUMBER: 60/017,699
? FILING DATE: 23-MAY-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Prestia, Paul F
? REGISTRATION NUMBER: 23,031
? REFERENCE/DOCKET NUMBER: ATG-50009-1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 610-407-0700
? TELEFAX: 610-407-0700
? TELEX: 846169
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 489 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-09-249-200-7

Query Match          12.6%   Score 187.5; DB 4; Length 489;
Best Local Similarity 40.8%; Pred.No. 9.9e-11;
Matches 40; Conservative      8; Mismatches 29; Indels 21; Gaps

QY    46 PPKGGDGEKGDPGEGKHGVGMGRGKIGKEL-----GDMDG 84
       ||| : |||| |::|||::||| |||| 
Db     141 PPAAEKGEGAAGRDGTPTVGQGPPSGKEAQLGLTAPEKOGATCAPCPNGS 200
              ||| |||| |::||| |||| 

QY    95 RNICTKTPTGKKDKGDKBKGLLGPCKGKAGTCVDCG 122
       |:|| |||| |::||| |::||| |||| 
Db     201 KGDILGTPKGEHGTKGDKDGLPLGNKGDMGMKGDG 238

RESULT          9
US-08-392-367B-2
? Sequence 2, Application US/08392367B
? Patent No. 5691197
? GENERAL INFORMATION:
? APPLICANT: Tytggevason, Karl
? APPLICANT: Elcomaa, Outi
? APPLICANT: Kangas, Maarit
? TITLE OF INVENTION: An Insolated DNA Sequence For a
? Patent No. 5691197
? TITLE OF INVENTION: No. 5691197el Macrophage Receptor with
? TITLE OF INVENTION: a Collagenous Domain and the
? TITLE OF INVENTION: Polypeptide Chain Encoded by
? TITLE OF INVENTION: such a Sequence
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fay, Sharpe, Beall, Fagan,
? ADDRESSEE: Minnich & McKee
? STREET: 1100 Superior Avenue
? STREET: Suite 700
? CITY: Cleveland
? STATE: Ohio
? COUNTRY: U.S.A.
? ZIP: 44114-2518
COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette, 3.50 inch,
? MEDIUM TYPE: 720 KB etc-able
```

COMPUTER: IBM PS/2, Model 35 SX  
OPERATING SYSTEM: DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,367B  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Minnich, Richard J.  
REGISTRATION NUMBER: 24,175  
REFERENCE/DOCKET NUMBER: TRV 2 009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (216) 861-5582  
TELEFAX: (216) 241-1666  
TELEX: (216) 980162  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 518 amino acids  
TYPE: amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-392-367B-2

Query Match 12.6%; Score 187.5; DB 1; Length 518;  
Best Local Similarity 40.8%; Pred. No. 1.1e-10;  
Matches 40; Conservative 8; Mismatches 29; Indels 21; Caps 1;

QY 46 PGPKGDCGKGPGEKGKGVGMGPKGIKGL-----GDMGD 84  
DB 170 PGPAAEKGKAGRGDTGPGVQPGPGSGKGLGLTGAFGKQATGAPGPRGKGS 229  
QY 85 RGNIGKTGPIGKKGKGLLGPGEKGKAGTVDCDG 122  
DB 230 KDGILGTPGKGEHTKGDGLGPNKNGDMKMGDTG 267

RESULT 10  
US-08-893-467A-2  
Sequence 2, Application US/08893467A  
Patent No. 6063901  
GENERAL INFORMATION:  
APPLICANT: Tryggvason, Karl  
APPLICANT: Elomaa, Outi  
APPLICANT: Kangas, Maarit  
TITLE OF INVENTION: An Insolated DNA Sequence For a  
Patent No. 6063901  
TITLE OF INVENTION: No. 5063901el Macrophage Receptor with  
TITLE OF INVENTION: a Collagenous Domain and the  
TITLE OF INVENTION: Polypeptide Chain Encoded by  
TITLE OF INVENTION: such a Sequence  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fay, Sharpe, Beall, Fagan,  
ADDRESSEE: Minnich & McKee  
STREET: 1100 Superior Avenue  
STREET: Suite 700  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: U.S.A.  
ZIP: 44114-2518  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch,  
MEDIUM TYPE: 720 Kb storable  
COMPUTER: IBM PS/2, Model 35 SX  
OPERATING SYSTEM: DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/893,467A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Minnich, Richard J.

REGISTRATION NUMBER: 24,175  
REFERENCE/DOCKET NUMBER: TRV 2 009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (216) 861-5582  
TELEFAX: (216) 241-1666  
TELEX: (216) 980162  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 518 amino acids  
TYPE: amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-893-467A-2

Query Match 12.6%; Score 187.5; DB 3; Length 518;  
Best Local Similarity 40.8%; Pred. No. 1.1e-10;  
Matches 40; Conservative 8; Mismatches 29; Indels 21; Gaps 1;

QY 46 PGPKGDCGKGPGEKGKGVGMGPKGIKGL-----GDMGD 84  
DB 170 PGPAAEKGKAGRGDTGPGVQPGPGSGKGLGLTGAFGKQATGAPGPRGKGS 229  
QY 85 RGNIGKTGPIGKKGKGLLGPGEKGKAGTVDCDG 122  
DB 230 KDGILGTPGKGEHTKGDGLGPNKNGDMKMGDTG 267

RESULT 11  
US-09-011-735-6  
Sequence 6, Application US/09011735B  
Patent No. 6110708  
GENERAL INFORMATION:  
APPLICANT: Wakamiya, No. 6110708utaka  
TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof  
FILE REFERENCE: 19036/34548  
CURRENT APPLICATION NUMBER: US/09/011,735B  
CURRENT FILING DATE: 1998-05-22  
EARLIER APPLICATION NUMBER: JP 7-209698  
EARLIER FILING DATE: 1995-08-17  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 161  
TYPE: PRT  
ORGANISM: Bovine  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (2)  
OTHER INFORMATION: Xaa is a protein-constituting amino acid  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (3)  
OTHER INFORMATION: Xaa is a protein-constituting amino acid  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (5)  
OTHER INFORMATION: Xaa is a protein-constituting amino acid  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (6)  
OTHER INFORMATION: Xaa is a protein-constituting amino acid  
US-09-011-735-6

Query Match 12.6%; Score 187; DB 3; Length 161;  
Best Local Similarity 29.7%; Pred. No. 2.4e-11;  
Matches 47; Conservative 28; Mismatches 69; Indels 14; Gaps 6;

QY 117 TVCDGGRYKFKVGLDLSIAKLKTSKMFVNVTAGIRETEKFFYIVQEEKNYRESLTHC 176  
DB 16 TILD-GHLRFQN-----AFSQYKKAIVFPDQAVG-----EKIFKTAGVKSYSQAQLC 65

QY 177 RIRGMLAMPKDEAANTLIADYVAKSGFRVFIQVNDLEREQYMTNTPLQYNSWNE 236  
 Db 66 REAKGGLASPSREANAVQWY-RAQKNAYLSNNDISTEGRTYPTGTEILV-YSNWAD 123  
 QY 237 GEP--SDPYGHEDCVEMLSGSRWNTDECHLIMYEVCEE 272  
 Db 124 GEPNDEQPCNVETIFPGKKNVPCSKOLLVICEF 161

RESULT 12  
 US-09-227-357-190  
 ; Sequence 190, Application US/05227357  
 ; Patent No. 6342581  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fischer et al.  
 ; TITLE OF INVENTION: 123 Human Secreted Proteins  
 ; FILE REFERENCE: P2010P1  
 ; CURRENT APPLICATION NUMBER: US/09/227,357  
 ; CURRENT FILING DATE: 1999-01-08  
 ; EARLIER APPLICATION NUMBER: PCT/US98/13684

; EARLIER FILING DATE: 1998-07-07  
 ; EARLIER APPLICATION NUMBER: 60/051,926  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/052,793  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,925  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,929  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/052,803  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/052,732  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,931  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,932  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,916  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,930  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,918  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,920  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/052,733  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/052,795  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,919  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/051,928  
 ; EARLIER FILING DATE: 1997-07-08  
 ; EARLIER APPLICATION NUMBER: 60/055,722  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,723  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,948  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,949  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,953  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,950  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,947  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,964  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/056,360  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,684  
 ; EARLIER FILING DATE: 1997-08-18

; EARLIER APPLICATION NUMBER: 60/055,984  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/055,954  
 ; EARLIER FILING DATE: 1997-08-18  
 ; EARLIER APPLICATION NUMBER: 60/058,785  
 ; EARLIER FILING DATE: 1997-09-12  
 ; EARLIER APPLICATION NUMBER: 60/058,664  
 ; EARLIER FILING DATE: 1997-09-12  
 ; EARLIER APPLICATION NUMBER: 60/058,660  
 ; EARLIER FILING DATE: 1997-09-12  
 ; EARLIER APPLICATION NUMBER: 60/058,661  
 ; EARLIER FILING DATE: 1997-09-12  
 ; NUMBER OF SEQ ID NOS: 672  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 190  
 ; LENGTH: 128  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (127)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (128)  
 ; OTHER INFORMATION: Xaa equals stop translation  
 ; US-09-227-357-190

Query Match 12.28; Score 181.5; DB 4; Length 128;  
 Best Local Similarity 52.98; Pred. No. 6.4e-11;  
 Matches 36; Conservative 8; Mismatches 21; Indels 3; Gaps 1;  
 QY 46 PGPKGDDGEGDGPGEKGKVGKVGKIGKELGMDGRGNIGKTPICKKDKGKGL 105  
 Db 61 PGRDGRGRGKGEKGTAGLNGKTPGLSLAGE---KDDQGETGKKGPICPEKGEVGP 117  
 QY 106 LGIPGK 113  
 Db 118 IGPPGPKG 125

RESULT 13  
 5514582-17  
 ; Patent No. 5514582  
 ; APPLICANT: CAPON, DANIEL J.; LASKY, LAURENCE A.  
 ; TITLE OF INVENTION: RECOMBINANT DNA ENCODING HYBRID  
 ; IMMUNOGLOBULINS  
 ; NUMBER OF SEQUENCES: 43  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/185,670  
 ; FILING DATE: 21-JAN-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 986,931  
 ; FILING DATE: 08-DEC-1992  
 ; APPLICATION NUMBER: 808,122  
 ; FILING DATE: 16-DEC-1991  
 ; APPLICATION NUMBER: 440,625  
 ; FILING DATE: 22-NOV-1989  
 ; APPLICATION NUMBER: 315,015  
 ; FILING DATE: 23-FEB-1989  
 ; SEQ ID NO:17:  
 ; LENGTH: 107  
 ; 5514582-17

Query Match 12.18; Score 180; DB 6; Length 107;  
 Best Local Similarity 35.48; Pred. No. 7.1e-11;  
 Matches 40; Conservative 20; Mismatches 47; Indels 6; Gaps 4;  
 QY 158 KFYIVQVEKNVRESLTHCRIRGGMAMPKDEAANTLIADYVAKSGFRVFIQVNDLERE 217  
 Db 1 KFFVTNHERPFSKVKALCSLRLGTVAIPRAENKAIOE-VAKTS---AFLGTDVTE 56

; SEQ ID NO 10

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2002, 12:34:21 ; Search time 42.99 Seconds  
(without alignments)  
1114.668 Million cell updates/sec

Title: US-09-600-932-2  
Perfect score: 1484  
Sequence: 1 MNGFASLLRNQFILLVFL.....NUTECHLMYFVCEFTKKK 277

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 17994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPREMBL19.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1484	100.0	277	4 Q9Y6Z7	Q9Y6Z7 homo sapien
2	728	49.1	272	11 Q9DC75	Q9DC75 mus musculus
3	716.5	48.3	271	4 Q9BWP8	Q9BWP8 homo sapien
4	338.5	22.8	378	6 Q9NIX4	Q9NIX4 sus scrofa
5	337	22.7	254	13 Q987A4	Q987A4 gallus gall
6	334.5	22.5	251	13 Q91908	Q91908 brachydanio
7	327.5	22.1	238	13 Q57451	Q57451 gallus gall
8	325	21.9	335	6 Q97748	Q97748 bos taurus
9	316	21.3	256	13 Q91909	Q91909 cyprinus ca
10	307	20.7	246	13 Q91907	Q91907 carassius a
11	297.5	20.0	236	6 Q28518	Q28518 macaca mula
12	296	19.9	248	6 Q9TT06	Q9TT06 ovis aries
13	294	19.8	248	6 Q9N0G1	Q9N0G1 equus cabal
14	291	19.6	245	6 Q28517	Q28517 macaca mula
15	289.5	19.5	248	4 Q96KE4	Q96KE4 homo sapien
16	289	19.5	248	6 Q95L88	Q95L88 equus cabal

17	288.5	19.4	248	4 Q96TE7	Q96TE7 homo sapien
18	282.5	19.0	224	11 Q9Z294	Q9Z294 rattus sp.
19	282.5	19.0	248	4 Q96TE8	Q96TE8 homo sapien
20	281.5	19.0	248	4 Q96TE9	Q96TE9 homo sapien
21	273	18.4	742	4 Q9BYH7	Q9BYH7 homo sapien
22	272.5	18.4	248	11 Q9C011	Q9C011 mus musculus
23	256	17.3	240	6 Q9XSK3	Q9XSK3 sus scrofa
24	211	14.2	268	2 Q9F685	Q9F685 streptococc
25	207	13.9	689	4 Q14055	Q14055 homo sapien
26	206.5	13.9	222	13 Q90XB2	Q90XB2 gallus gall
27	205	13.8	645	9 Q9XJB9	Q9XJB9 escherichia
28	203	13.7	688	11 Q07643	Q07643 bacterioph
29	203	13.7	1723	5 Q9GQ31	Q9GQ31 mus musculus
30	198.5	13.4	479	2 Q9LAC2	Q9LAC2 hydra atten
31	196.5	13.2	289	4 Q9HJ2	Q9HJ2 escherichia
32	195.5	13.2	1779	5 Q9VMV4	Q9VMV4 drosophila
33	195.5	13.2	465	2 Q9AGC4	Q9AGC4 streptococc
34	193.5	13.0	291	2 Q9F690	Q9F690 streptococc
35	193	12.9	1366	4 Q15177	Q15177 homo sapien
36	191.5	12.8	410	2 Q9F691	Q9F691 streptococc
37	190.5	12.8	380	2 Q9F692	Q9F692 streptococc
38	189.5	12.8	358	5 Q23222	Q23222 caenorhabdi
39	189	12.7	397	2 Q9F686	Q9F686 streptococc
40	189	12.7	404	6 Q95J96	Q95J96 macaca mula
41	189	12.7	2944	11 Q63870	Q63870 mus musculus
42	188	12.7	291	5 Q23422	Q23422 caenorhabdi
43	188	12.7	310	13 Q90612	Q90612 gallus gall

ALIGNMENTS

RESULT 1  
Q9Y6Z7 PRELIMINARY; PRT; 277 AA.  
ID AC Q9Y6Z7;  
DT 01-NOV-1999 (TRENBLREL. 12, Created)  
DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)  
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)  
DE COLLECTIN 34.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9240769; PubMed=10224141;  
RA Ohtani K., Suzuki Y., Eda S., Kawai T., Kase T., Yamazaki H.,  
RA Keshi H., Sakai Y., Fukuchi A., Sakamoto T., Wakamiya N.;  
RT "Molecular cloning of a novel collectin from liver (CL-L1).";  
RL J. Biol. Chem. 274:13681-13689(1999).  
DR EMBL; AB002631; BAA81747.1; -;  
DR HSSP; P20693; IHLJ.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR001304; lectin\_c.  
DR Pfam; PF01391; Collagen; 1.  
DR Pfam; PF00059; lectin\_c; 1.  
DR SMART; SM00034; CLECT; 1.  
DR PROSITE; PS00615; C\_TYPE\_LLECTIN\_1; 1.  
DR PROSITE; PS50041; C\_TYPE\_LLECTIN\_2; 1.  
SQ SEQUENCE 277 AA; 30733 MW; 9736861CEBDC5C25 CRC64;

Query Match 100.0%; Score 1484; DB 4; Length 277;  
Best Local Similarity 100.0%; Pred. No. 1.6e-128;  
Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNGFASLLRNQFILLVFLVFLQISGLGIDSRPTAEVCATHTISPGKGDGKGDGE 60  
Db 1 MNGFASLLRNQFILLVFLVFLQISGLGIDSRPTAEVCATHTISPGKGDGKGDGE 60

QY	61	EGKHGKVGVRMGPKIGELGDMGDGNIGKGTPTGKKGDKGKGLLIGIPGKAGTVD	120
Db	61	EGKHGKVGVRMGPKIGELGDMGDGNIGKGTPTGKKGDKGKGLLIGIPGKAGTVD	120
QY	121	CGRYKFKVGQLDISARLKTSMKFKVKNVIAGTETKEFKFYIIVOEKYNRESLTHCRIG	180
Db	121	CGRYKFKVGQLDISARLKTSMKFKVKNVIAGTETKEFKFYIIVOEKYNRESLTHCRIG	180
QY	181	GMLAMPKDEAANTLIADIVYAKSGFFRVFTGVNDLEREGQYTFDNTPTONTYNNWGEPS	240
Db	181	GMLAMPKDEAANTLIADIVYAKSGFFRVFTGVNDLEREGQYTFDNTPTONTYNNWGEPS	240
QY	241	DPYGHEDCVEMLSGSGRWNDTECHLTWYFYCEFIKKK	277
Db	241	DPYGHEDCVEMLSGSGRWNDTECHLTWYFYCEFIKKK	277

RESULT 2

Q9DC75

ID

Q9DC75

PRELIMINARY;

PRT; 272 AA.

AD

Q9DC75

Q9DC75

Q9DC75

DT

01-JUN-2001

(rEMBLrel. 17, Created)

DT

01-JUN-2001

(rEMBLrel. 17, Last sequence update)

DT

01-DEC-2001

(rEMBLrel. 19, Last annotation update)

DE

1010001H16RIK

PROTEIN.

GN

1010001H16RIK

OS

Mus musculus (Mouse).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

SEQUENCE FROM N.A.

RC

STRAIN=C57BL/6J; TISSUE=HEART;

RX

MEDLINE=21085660; PubMed=11217851;

RA

Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA

Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA

Aizawa K., Matsuda H., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA

Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,

RA

Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA

Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,

RA

Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA

Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA

Salai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA

Brake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA

Brownstein M.J., Bult C., Fietcher C., Fujita M., Gariboldi M.,

RA

Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA

Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA

Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA

Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA

Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,

RA

Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,

RA

Hayashizaki Y.;

RT

"Functional annotation of a full-length mouse cDNA collection."

RL

Nature 409:685-690(2001).

DR

EMBL; AK003121; BAB22581.1; -.

DR

HSSP; P20693; 1HL7.

DR

MGI; MGI:1918943; 1010001H16RIK.

DR

InterPro; IPR000087; Collagen.

DR

InterPro; IPR001304; lectin\_c.

DR

Pfam; PF01391; Collagen; 1.

DR

Pfam; PF00059; lectin\_c; 1.

DR

SMART; SM00034; CLECT; 1.

DR

PROSITE; PS00615; C-TYPE-LECTIN\_1; 1.

DR

PROSITE; PS50041; C-TYPE-LECTIN\_2; 1.

SQ

SEQUENCE 272 AA; 28975 MW; DEC471493CD16B95 CRC64;

Query Match

49.1%; Score 728; DB 11; Length 272;

Best Local Similarity

48.3%; Pred. No. 55-59;

Matches 128; Conservative 56; Mismatches 73; Indels 8; Gaps

QY 14

ILLVIFLLIQSLGLDIDSRPTAEVCAHTTSPGKGDEGKVGPKGKAGTVD

120

Db 14

ILLVIFLLIQSLGLDIDSRPTAEVCAHTTSPGKGDEGKVGPKGKAGTVD

120



```

RESULT 4
Q9N1X4 PRELIMINARY; PRT; 378 AA.
AC Q9N1X4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LUNG SURFACTANT PROTEIN D PRECURSOR.
GN SFTPD.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP MEDLINE=20109098; PubMed=10640760;
RA van Eijk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
RA Lawson P.R.;
RT "Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
RT localisation and tissue distribution.";
RL J. Immunol. 164:1442-1450(2000).
DR EMBL; AF132496; AAF22145.2; -
DR HSSP; P35247; 1908.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 378 LUNG SURFACTANT PROTEIN D.
FT SEQUENCE 378 AA; 37986 MW; 3504E8C1E56C341D CRC64;
SQ
Query Match 22.8%; Score 338.5; DB 6; Length 378;
Best Local Similarity 32.8%; Pred. No. 5.7e-23;
Matches 81; Conservative 35; Mismatches 110; Indels 21; Gaps 6;
QY 45 SPGKGGDGGKPGEGKHGKVRMGPKIGELGDMGDR---GNIGTKGPIGKKGDKG 101
DB 134 TPFGKGTGKPGVGLGALGQSGTGARGPGLGKRGAPGAPGAGPAGATGPGQ 193
QY 102 EKLGLIPGEGKAGTVCDG-----RYKFGQLDISIARLKTSMKFKVNV-- 148
DB 194 PSARGPPGLKDRGPPGERGAKGESGLPGLALRQQVETLQGVQLQKAFSQYKVEL 253
QY 149 IAGIRETEEFYIVQEKKNYRESLTHCRIRGMLAMPKDEAANTLIADYVAKSGFFRVF 208
DB 254 PFNGRGVGEKIFKGTGFEKTDQAQVCTQAGQMASPRSETENEALSQLVTAQN-KA 312
QY 209 IGVDLEREGQYMTDNTPLQYNNWGEPSD---PYGHEDCVEMLSGRWNTDTECHLT 265
DB 313 LSKTDIKENFYPTGPELV-YANWAPGEPNNGSSGAENCVEIFPNKGWNRKAGCEL 371
QY 266 MYFVCEP 272
DB 372 RVICEF 378
RESULT 5
Q98TA4 PRELIMINARY; PRT; 254 AA.
AC Q98TA4
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE MANNOSIDE-BINDING LECTIN PROTEIN PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

```

```

OC Gallus
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjoldt K.;
RT "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.";
RL Immunogenetics 51:955-964(2000).
DR EMBL; AF231714; AAK30298.1; -
DR HSSP; P19999; 1YTT.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000005; HTHARAC.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE-LECTIN_1; UNKNOWN_1.
DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
KW Signal; Lectin.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 254 MANNOSIDE-BINDING LECTIN PROTEIN.
FT VARIANT 234 234 D -> V.
FT SEQUENCE 254 AA; 27376 MW; C924428643441AED CRC64;
SQ
Query Match 22.7%; Score 337; DB 13; Length 254;
Best Local Similarity 31.8%; Pred. No. 4.7e-23;
Matches 92; Conservative 54; Mismatches 83; Indels 60; Gaps 15;
QY 6 SLLRNQFTLLVFLQLQSLGIDIDSRPAEV-----CATHIS--PGPKGD---G 53
DB 2 TLQPSFALLCLSLMATSL-LTTD-KPEKMYSCPIIQCSAPAVNGLPGRDGRGPKG 59
QY 54 EKDGPGE--EGKHGKVRMGPKIGELGDMGDRGNIGTKGPIGKKGDKGKGL----- 106
DB 60 EKDGPGELRGLGKAGPGLGKEV-----GPOGEGKGGKGGVIVTDDL 107
QY 107 --GIPGKAGTV-CDGGRYKRVGQLDISIARLKTSMKFKVNVYAGIRETEEFYIV 163
DB 108 HRQITDEAKIRVLEDDLRYKKAL-----SLKDVNV-----GKKMFVST 148
QY 164 QBEKNYRESLTHCRIRGMLAMPKDEAANTLIADYVAKSGFFRVFVGVNDRERGGYMT 223
DB 149 GKYNFEKSKLCAKAGSVLASPRNEANTALKOLIDPSS--QAYIGISDAQTEGRFMYL 206
QY 224 DNTPLQYNNWGEPSDPTGHDCEVEMLSGRWNTDTECHLTWYFVCE 271
DB 207 SGGPL-TYSNWKPGEPNN-HKNECDCAVDSGKNWDLDCSNINFIICE 253
RESULT 6
Q91908 PRELIMINARY; PRT; 251 AA.
AC Q91908
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MANNOSIDE-BINDING-LIKE LECTIN PRECURSOR.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjoldt K.;
RT "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.";

```

```

RL Immunogenetics 51:955-964(2000).
DR EMBL: AF227738; AAF63469.1; -.
DR HSP; P19999; IAFB.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 1.
DR DR PF00059; Lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE; PS00441; C_TYPE_LECTIN_2; 1.
KW SIGNAL; LECTIN.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 251 MANNOSE BINDING-LIKE LECTIN.
FT VARIANT 21 21 M -> L.
FT VARIANT 204 204 K -> N.
SQ SEQUENCE 251 AA; 26829 MW; 12D0ABD06B6E3B11 CRC64;

Query Match 22.5%; Score 334.5; DB 13; Length 251;
Best Local Similarity 30.4%; Pred. No. 7.9e-23;
Matches 79; Conservative 49; Mismatches 107; Indels 25; Gaps 8;

QY 14 ILLVFLIQISGLDIDSRPTAEVCAHTISPGPKGDDGEGKPGEGKHGKVGGMGPX 73
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 74 GIKGELGMDGRNIGKGTGKKGKGEKGLGIPGEGKAGTVCDCGGRYKRVGQLDI 133
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 134 STARLKTSMKFKVNI--AGIRETEKFFYIVQBEKNYRESLTHCRIRGMLAMPKDEAA 191
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 110 ELQKLSKIALIEKVNVKTKKVGQKYVTDVEETDKGMOYCSNGGALVLPRTLEE 169
QY 192 NTLIADYVAKSGFRFVIGVNDLEREGQYMTDNTPLQNSNNWEGPSDPYGHEDCVEM 251
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 252 LSSGRWNTDCHLTLYFFVCE 271
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 228 ADSGLMDVDSOSLYPIICE 247
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 7
O57451 PRELIMINARY; PRT; 238 AA.
AC O57451;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE MANNAN-BINDING LECTIN (FRAGMENT).
GN CMBL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN; TISSUE=LIVER;
RA Laursen S.B.;
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF022262; BAB94071.1; -.
DR HSP; P19999; 1YT.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE; PS00441; C_TYPE_LECTIN_2; 1.
KW LECTIN.
FT NON_TER 238 238
SQ SEQUENCE 238 AA; 25645 MW; E5C9B5197AAE64E3 CRC64;

```

```

Query Match 22.1%; Score 327.5; DB 13; Length 238;
Best Local Similarity 32.1%; Pred. No. 3.3e-22;
Matches 80; Conservative 48; Mismatches 70; Indels 51; Gaps 12;

QY 39 CAHTIS--PGFKGDD--GEKDPGE--EGKHGKVGGMGPKGKGLGMDGMDGRNIGKT 91
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 24 CSPAVNGLPGSRDGDGEGKPGEGRLGQLGPKGAGPQLKGEV----- 71
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 92 GPTEKKGDKGKGLL-----GIPGEGKAGTV--CDGGRYKRVGQGLDLSIARLKTSMK 143
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 72 GPQGEKGQGERGIVITDHLRQITDLEAKIRVLEDDLSRYKKAL-----SLK 119
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 144 EVKNVIAGIRETEKFFYIVQBEKNYRESLTHCRIRGMLAMPKDEAAATLIADYVAKSG 203
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 120 DVNI-----GKMFVSTGKKYNEFKGSLCAKAGSVLASPRNEAENTALKDLIDPS 172
QY 204 FFRVFVIGVNDLEREGQYMTDNTPLQNSNNWEGPSDPYGHEDCVEMLSGRWNTDEC- 262
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 173 --QAYGISDAQTEGREGWYLSGGL--TYSNNKPGEPNN--HKNECAVIEDSGKNOLDSCS 228
QY 263 HLWYFVCE 271
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 229 NSNIFICE 237
Db :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 8
O57748 PRELIMINARY; PRT; 335 AA.
AC O57748;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CONGLUTININ.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93277452; PubMed=7684896;
RA Lu J., Laursen S., Thiel S., Jensenius J., Reid K.;
RT "The cDNA cloning of conglutinin and identification of liver as a
primary site of synthesis of conglutinin in members of the Bovidae.";
RL Biochem. J. 292:157-162(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=94128104; PubMed=8297370;
RA Kawasaki N., Itoh N., Kawasaki T.;
RT "Gene organization and 5'-flanking region sequence of conglutinin: a
C-type mammalian lectin containing a collagen-like domain.";
RL Biochem. Biophys. Res. Commun. 198:597-604(1994).
DR EMBL: D25302; BAA04983.1; -.
DR EMBL: D25296; BAA04983.1; JOINED.
DR EMBL: D25297; BAA04983.1; JOINED.
DR EMBL: D25299; BAA04983.1; JOINED.
DR EMBL: D25300; BAA04983.1; JOINED.
DR EMBL: D25301; BAA04983.1; JOINED.
DR HSP; P35247; IB08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 2.

```

```

DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS00041; C_TYPE_LLECTIN_2; 1.
FT CONFLICT 236 236 E -> V (IN REF. 2).
SQ SEQUENCE 335 AA; 34702 MW; 39D3A30BC76C134C CRC64;

Query Match 21.9%; Score 325; DB 6; Length 335;
Best Local Similarity 33.8%; Pred. No. 8.6e-22;
Matches 80; Conservative 33; Mismatches 96; Indels 28; Gaps 8;

QY 46 PGPFGDCEKGDGEGSKGKVGKMGKPGKIGKELGDMGDRGNIGKTPGKKGKGLGIPGKAGTVCDCGRYKFKVGMCPK 105
Db 117 PGPGLKGEKAGPAGGAPGAGVTPGSA-----IGPQPGSARGPGLAGDRGDGPE 170
QY 106 LGIPGKAGKAG-----TVCDCGRYKFKVGMGLDISIARLKTSMKFKVKNVITAGIRETEE 157
Db 171 TGAKGESLAENVNALKQRVTILD-CHLRFTQN-----AFSQYKKAIVLPDQAVG-----E 220
QY 158 KFYIVQBEKNYRESLTHCRINGMLAMPKDEAANTLIADYVAKSGFRRVFGVNDLRE 217
Db 221 KIFKTAGAVKSYDAEQLCREAKQGLASPRSSAENEAQVQMV-RAQEKNAVLSMNDIISTE 279
QY 218 GQYMTDNTPLQVSNWNGEP--SDPYGHEDCVEMLSGGRWNTDTECHLTMFVCEFP 272
Db 280 GRFYPTGEILV-YSNWADGEPNNSDEGQPCNVEIFPDGKNDVPCSKQLLVICGF 335

RESULT 9
Q91909
ID Q91909 PRELIMINARY; PRT; 256 AA.
AC Q91909;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MANNOSIDE BINDING-LIKE LECTIN PRECURSOR.
GN MBL.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprinus.
OX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjold K.;
RT "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.";
RL Immunogenetics 51:955-964(2000).
DR EMBL: AF227737; AAF63468.1; -.
DR HSSP: P35247; 1B08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR PROSITE: PS00041; C_TYPE_LLECTIN_2; 1.
KW Signal; Lectin.
FT SIGNAL 1 23 POTENTIAL.
FT VARIANT 24 256 MANNOSIDE BINDING-LIKE LECTIN.
FT CHAIN 235 235 S -> T.
SQ SEQUENCE 256 AA; 26934 MW; D019291D1167730D CRC64;

Query Match 21.3%; Score 316; DB 13; Length 256;
Best Local Similarity 31.6%; Pred. No. 4.1e-21;
Matches 84; Conservative 39; Mismatches 119; Indels 24; Gaps 8;

QY 14 ILLVFLQLQISGLDIDSRTAEVCATHTISPGKGDGEGKGGKGVGMCPK 73
Db 1 LLLVFLQALQL-----LD-GAEPONLNCPAY---GGVPGTPGHNLPGDRGRDGDGALGPK 52
QY 74 GIKGELGDMGDRGNIGKTPGKKGKGLGIPGKAGTVCDCGRYKFKVGMCPK 133
Db 53 GEKGESG-VSVQGPCKAGPCTAGKGGSGGQSGPGS-----ESVLESLS 101
QY 134 STARLKTSMKFKVKNV--IAGIRETEKFFYIVQBEKNYRESLTHCRINGMLAMPKDEA 191
Db 102 ETQQLAKIATEKFSVSVCHFRKVGQKYITDGVVGNFDGLKSCMEFGSTWVSPRTSAE 161

Db 11 LLLVFLQALQL-----LD-GAEPONLNCPAY---GGVPGTPGHNLPGDRGRDGDGALGPK 62
QY 74 GIKGELGDMGDRGNIGKTPGKKGKGLGIPGKAGTVCDCGRYKFKVGMCPK 133
Db 63 GEKGESG-VSVQGPCKAGPCTAGKGGSGGQSGPGS-----ESVLESLS 111
QY 134 STARLKTSMKFKVKNV--IAGIRETEKFFYIVQBEKNYRESLTHCRINGMLAMPKDEA 191
Db 112 ETQQLAKIATEKFSVSVCHFRKVGQKYITDGVVGNFDGLKSCMEFGSTWVSPRTSAE 171
QY 192 NTLIADYVAKSGFF--RVFVGVNDLRECGYMTDNTPLQVSNWNGEPSPDPYGHEDCV 249
Db 172 NQALLKLVVSSGLSSKKPYIGVTDRETEGRFVNTGKQL-TFTNMGCPQPDYKGLQDCG 230
QY 250 EMLSSGRWNTDTECHLTMFVCEFIKK 275
Db 231 VIEDSGLWDDGSCGDIRPIMCEIDNK 256

RESULT 10
Q91907
ID Q91907 PRELIMINARY; PRT; 246 AA.
AC Q91907;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE MANNOSIDE BINDING-LIKE LECTIN PRECURSOR (FRAGMENT).
GN MBL.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=20456722; PubMed=11003389;
RA Vitved L., Holmskov U., Koch C., Teisner B., Hansen S., Skjold K.;
RT "The homologue of mannose-binding lectin in the carp family Cyprinidae
RT is expressed at high level in spleen, and the deduced primary
RT structure predicts affinity for galactose.";
RL Immunogenetics 51:955-964(2000).
DR EMBL: AF227739; AAF63470.1; -.
DR HSSP: P35247; 1B08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; UNKNOWN_1.
DR PROSITE: PS00041; C_TYPE_LLECTIN_2; 1.
KW Signal; Lectin.
FT SIGNAL 1 13 POTENTIAL.
FT VARIANT 145 145 S -> F.
SQ SEQUENCE 246 AA; 25709 MW; AB692282D289D0D5 CRC64;

Query Match 20.7%; Score 307; DB 13; Length 246;
Best Local Similarity 31.7%; Pred. No. 2.6e-20;
Matches 83; Conservative 39; Mismatches 116; Indels 24; Gaps 8;

QY 14 ILLVFLQLQISGLDIDSRTAEVCATHTISPGKGDGEGKGGKGVGMCPK 73
Db 1 LLLVFLQALQL-----LD-GAEPONLNCPAY---GGVPGTPGHNLPGDRGRDGDGALGPK 52
QY 74 GIKGELGDMGDRGNIGKTPGKKGKGLGIPGKAGTVCDCGRYKFKVGMCPK 133
Db 53 GEKGESG-VSVQGPCKAGPCTAGKGGSGGQSGPGS-----ESVLESLS 101
QY 134 STARLKTSMKFKVKNV--IAGIRETEKFFYIVQBEKNYRESLTHCRINGMLAMPKDEA 191
Db 102 ETQQLAKIATEKFSVSVCHFRKVGQKYITDGVVGNFDGLKSCMEFGSTWVSPRTSAE 161

```

```
QY 192 NTLIADYVAKSGF--FRVFIGVNDLEREGQYMTNTPLQYSNWNEGSPDPYGHEDCV 249
Db 162 NOALIKLVSSGLSKKPYIGVTRKTEGQFVDETKQL-TFTNWGPGQDDYKGLQDCG 220
QY 250 EMLSSGRWNTECHLTMFVCE 271
Db 221 VIEDTGLWDGGGDIRPME 242

RESULT 11
Q28518
ID Q28518 PRELIMINARY; PRT; 236 AA.
AC Q28518;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE MANNOSE-BINDING PROTEIN C (FRAGMENT).
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Medges T., Ota T., Fauber A.I., Sastry K.N.;
RT "Characterization of two mannose-binding protein cDNAs from rhesus
RL monkey (Macaca mulatta): structure and evolutionary implications.";
DR EMBL: L43911; AAB48071.1;
DR HSSP: P11226; 1HUP.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
FT NON TER 1
SQ SEQUENCE 236 AA; 24911 MW; 449CBE887C89732E CRC64;

Query Match 20.0%; Score 297.5; DB 6; Length 236;
Best Local Similarity 32.0%; Pred. No. 1.9e-19;
Matches 78; Conservative 33; Mismatches 110; Indels 23; Gaps 7;

QY 38 VCATHISPGKDDGK-----GDPGEKKGKVGKMGKPGKIGKELGMDGRGNICKTGP 93
Db 5 ICTNSTGGPQHADGSPVACNSPGINGFPKDRDGTGKGEKGFQ--QGLRGLQGPCK 63
QY 94 IGKKDKGKGLLIGPEKKGKAGTVDC-----GRYKFGVGLDISIARLTKSMKTKVNV 148
Db 64 LGPFGNGSSSPGKQKGDGSPDCSSLAASERK---ALQTEMARIKKWLTFSLG- 119
QY 149 IAGIRETEKFFYIVQEKYRESLTHCRIRGGMAMPKDEAANTLIADYVAKSGFPRVF 208
Db 120 ----RQVGNKFLINGEMTFDKVKALCARQASVATPNAENAIQNLKE-----EAF 171
QY 209 IGVNLEREGQYMTNTPLQYSNWNEGSPDPYGHEDCVMLSSGRWNTECHLTMF 268
Db 172 LGITDTEGEFVLTGNKL-TVTWNDDGNAGSNEDCVLLKNGKWNIDPCSSSLA 230
QY 269 VCEF 272
Db 231 LCEF 234

RESULT 12
Q28518
ID Q28518 PRELIMINARY; PRT; 248 AA.
AC Q28518;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DE PULMONARY SURFACTANT PROTEIN A (PULMONARY SURFACTANT-ASSOCIATED
OS Protein A).
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Braams G.A., Yao L.-J., Inchley K., Brickenden A., Han V.K.M.,
RT "Ovine surfactant protein cDNAs: use in studies on fetal lung growth
RL and maturation after prolonged hypoxemia.";
DR EMBL: AF211856; AAF18995.1;
DR EMBL: AF076633; AAF31148.1;
DR HSSP: P19999; 1YTT.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 1.
DR SMART: SM00059; lectin_c; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
FT NON TER 1
SQ SEQUENCE 248 AA; 26394 MW; D65E7293BBFF1FD9 CRC64;

Query Match 19.9%; Score 296; DB 6; Length 248;
Best Local Similarity 31.7%; Pred. No. 2.7e-19;
Matches 85; Conservative 34; Mismatches 113; Indels 36; Gaps 10;

QY 16 LVLFLLAQISGLDIDSRPTAEVCATHTISPGKGDGKDGKGEKKGKVGKMGKGI 75
Db 6 LFLMLNVAAGLECD---TREVCLG---SPGIPGTPGSHGLPGRDGRGKDPGPPG 59
QY 76 KQELGDM-GDRGNIGKTPGKKGKGLIGPEKKGKAGTVDCGRYKRVGQGLDIS 134
Db 60 MGPPGPGPLPGRDGMTGAPGLPGERGEK---GEPGERGPPG-----FPAYLD-- 104
QY 135 IARLTKSMKTKVNV-----AGIRETEKFFYIVQEKYRESLTHCRIRGMLA 184
Db 105 --EELQGLTHIRHOVLQSQGVLLILOGSMLEVGKVFSTNGOSLNFDALKELCARAGHIA 163
QY 185 MPKDEAANTLIADYVAKSGFPRVFIGNVNDLEREGQYMTNTPLQYSNWNEGSPDPY 244
Db 164 APRSPNEEAITSVKKNTY-AVLGLAEGPTAGDYFYLDAVPV-NYTNWYGPGRG-RG 220
QY 245 HEDCVMLSSGRWNTECHLTMFVCE 272
Db 221 KEKCVETDGDNDKNCLOYLAIACEF 248

RESULT 13
Q28518
ID Q28518 PRELIMINARY; PRT; 248 AA.
AC Q28518;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE PULMONARY SURFACTANT PROTEIN A.
OS Equus caballus (Horse).
```

RC TISSUE=LIVER;  
RX MEDLINE=97031450; PubMed=877375;  
RA Moques T., Ota T., Tauber A.I., Sastry K.N.;  
RT "Characterization of two mannose-binding protein cDNAs from rhesus  
RT monkey (*Macaca mulatta*): structure and evolutionary implications.";  
RL Glycobiology 6:543-550(1996).  
DR EXBL: L43912; ARB48070.1; -.  
DR HSP; P19999; IYTT.  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR001304; lectin\_c.  
DR Pfam: PF00059; lectin\_c; 1.  
DR SMART: SM00034; lectin\_1.  
DR PROSITE: PS00615; C\_TYPE\_LECTIN\_1; 1.  
DR

```
Query Match      19.5%; Score 289.5; DB 4; Length 248;  
Best Local Similarity 31.1%; Pred. No. 1.1e-18;  
Matches       75; Conservative    37; Mismatches   100; Indels     29; Gaps        7:  
  
QY      34 PTAEVCAHTTIS--PGPKGDCEKDGECCKGKGKVGRMGPKGIKGELDMCDRGNTICKT 91  
         ||::|||:||::|||||::|||::|||::|||::|||::|||  
Db       33 PAVIACSSPGLNPPFGKDGRDTKTGEKEGPQ-----GLRIGQQPFKLGGPCGNPEPS 85  
         ||::|||:||::|||||::|||::|||::|||::|||::|||  
QY       92 GPITKKGDKGSKGLLTPGEEKAKGTCTDCGRYRFVGGOLDISTARLKTSMKFVKNNIAG 151  
         |||::|||::|||::|||::|||::|||::|||::|||::|||  
Db       86 GSFGPKQGQDPG-----KSFDGSLSLAASERK---ALOTEMARIKWLTFSLG---- 131  
         |||::|||::|||::|||::|||::|||::|||::|||::|||  
QY      152 IRTEEKFFYIVQEKNRYRESTHCRINGCNLAMPKDEAANTLIADYVAKSGFRRVFVG 211
```

```
Db 132 -KQVGNKFELTNGEINTFEKVKALCVKQASVATPRNAENGAIQNLKE-----EAFIGI 186
QY 212 NDLEREGQYMFDTNTPLONYSNWNEGEPSPYGHEDCVEMISSGRWNDECHLTWYFVCE 271
Db 187 TDEKTEGQFVDLTGNEL-TYTNWNEGEPNAGSDEDCVLLKNGQWNVPCSTSHLAVCE 245
QY 272 F 272
Db 246 F 246
```

Search completed: July 3, 2002, 12:40:21  
Job time: 360 sec



Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
WPI; 1999-551358/46.  
N-PSDB; AAZ33973.

New secreted and transmembrane polypeptides and their polynucleotides, useful for treating blood coagulation disorders, cancers and cellular adhesion disorders -

Claim 12; Fig 37; 530pp; English.

The present invention describes secreted and transmembrane polypeptides and their polynucleotides. The nucleotide sequences are useful as sources of probes, primers, for chromosome mapping, and for generation of antisense sequences. They can also be used to create transgenic animals. The proteins can be used to treat a variety of diseases and disorders, depending on their function. Diseases that may be treated include blood coagulation disorders, cancers and cellular adhesion disorders. They may also be used to raise antibodies. AAZ33891 to AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and polypeptide sequence given in the exemplification of the present invention.

Sequence 277 AA;

Query Match 99.2%; Score 1472; DB 20; Length 277;  
Best Local Similarity 99.3%; Pred. No. 7.8e-141;  
Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 MNGFASLLRRNQFTLLVFLQIQSLGDLDSRPTAEVCATHPTSPGKGGDGGKGGDGG 60  
1 mngfasllrrnqftllvflqigsldidsrptaevcathptspgkggdggkggdgg 60

61 EGKHGKVRGMPKGIKGEIGDMGDRGNIGTGPFGKGDGKEXGLGIPGKGGKAGTVCD 120  
61 egkhgkvrmpkgikgeigdmgdgngtgpfgkgdgkexglgipgkggkagtvcd 120

121 CGEYRKFGQLDISIARLTKSKKFKVNWVINGTRETKEFYFYVQEKNYRSLTHCRIRG 180  
121 cgyrkyfvgqldisiarlktmkfkvnviagireteekfyfyvgeeknyreslthcrirg 180

181 GMLAMPKDEAANTLIADYVAKSGFFRVFGVNDLEREGGYMTFDTNPQNTSNNEGEPS 240  
181 gmlampkdeaanliadyvaksgffrvfvgvndlereggymsctnclpqnysnnegeps 240

241 DPGHEDCEVEMLSGRWNDTECHTMFYCEFIKKK 277  
241 dpyghedcevmlsgrwndtechltmfvcefiykkk 277

RESULT	3
ID	AAB44254
ID	AAB44254 standard; Protein; 277 AA.
XX	AC AAB44254;
XX	08-FEB-2001 (first entry)
XX	Human PRO702 (UNQ366) protein sequence SEQ ID NO:97.
DE	Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW	expressed sequence tag; detection; cancer.
KW	Homo sapiens.
XX	OS
XX	WO2000053756-A2.
PN	PD
XX	14-SEP-2000.
XX	18-FEB-2000; 2000WO-US04341.
PFF	08-MAR-1999; 99WO-US05028.
XX	
PPR	

PR	12-MAR-1999;	99US-0123957.
PR	29-MAR-1999;	99US-0126773.
PR	21-APR-1999;	99US-0130232.
PR	28-APR-1999;	99US-0131445.
PR	14-MAY-1999;	99US-0134287.
PR	23-JUN-1999;	99US-0141037.
PR	26-JUL-1999;	99US-0145698.
PR	29-OCT-1999;	99US-0162506.
PR	30-NOV-1999;	99WO-US28313.
PR	04-DEC-1999;	99WO-US28551.
PR	02-DEC-1999;	99WO-US28565.
PR	16-DEC-1999;	99WO-US30095.
PR	30-DEC-1999;	99WO-US31243.
PR	30-DIC-1999;	99WO-US31274.
PR	05-JAN-2000;	2000WO-US00219.
PR	06-JAN-2000;	2000WO-US00277.
PR	06-JAN-2000;	2000WO-US00376.
XX		
XX	(GETH ) GENENTECH INC.	
XX		
PI	Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;	
PI	Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;	
PI	Goddard A, Godowski PJ, Grimaldi CJ, Guney AL, Hillan KJ;	
PI	Kajava LJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;	
PI	Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;	
XX		
XX	WPI; 2000-611443/58.	
DR	N-PSDB; AAC78480.	
XX		
PPT	Novel PRO polypeptides and polynucleotides used in detection methods,	
PPT	to target bioactive molecules to specific cells, and to modulate	
PPT	cellular activities -	
XX		
XX	Claim 12; Fig 37; 636pp; English.	
PS		
PS	AAC78458 to AAC78599 represent polynucleotide and EST (expressed	
XX	sequence tag) sequences which encode secreted or transmembrane PRO	
CC	polypeptides. The PRO polynucleotides and polypeptides have cytostatic	
CC	activity. The polynucleotides and polypeptides can be used for detecting	
CC	the presence of PRO polypeptides in samples, for linking bioactive	
CC	molecules to cells and for modulating biological activities of cells,	
CC	using the polypeptides for specific targeting. The polypeptide targeting	
CC	can be used to kill the target cells, e.g. for the treatment of cancers.	
CC	The polypeptide pairs provide specific targeting of bioactive molecules	
CC	to cells. AAC78600 to AAC78687 represent PCR primers and probes used in	
CC	the isolation of the PRO polynucleotide sequences.	
XX		
SQ	Sequence 277 AA;	
SQ		

Query Match	99.2%	Score 1472;	DB 21;	Length 277;
Best Local Similarity	99.3%;	Pred. No. 7.8e-141;		
Matches 275;	Conservative 1;	Mismatches 0;	Gaps 0;	
QY	1	MNGFASLLRRNQFILLVLIQLIQSLGELDISDPSPTAECATHISPGPKGDDGKGDPGE	60	
DB	1	mngfasllrrnqfllvllqigslgldidsptaevacathispgpkgdgkgdpge	60	
QY	61	EKGHKVGVRMPKGIXGELGDMGRGNIGKTGCTCKKGDKGKGLLIGPEKGRAGTVCD	120	
DB	61	egkhkgvrmpkxigxelgdmgrgnigktgctckkgdkgkgliligpekgagtvcd	120	
QY	121	CGRYRKFGQLDTSIARLTSKMFKNVNIAGTRETKEFYFVQVEKNYRSLCHRCIRG	180	
DB	121	cgryrkrfvgqldisarltskmfkvnviagireteekfyfveeknyreslthcrrig	180	
QY	181	GMLAMPKDEAANTLIADYVAKSGFFRFVIGVNDLREGQYMFNTPLQNTSNNEGEPS	240	
DB	181	gmlampkdeaanliadyvaksgffrfvignvndleregqymfntplqntsnnegeps	240	
QY	241	DPYGHEDCVMELSGRWNDDTECHTMTVCFEIKKKK	277	
DB	241	dpvqghedcvmelsgarwndtechtmtvcefikkkk	277	



07-DEC-1999	(first entry)
Human	PRO702 nucleotide sequence.
Human;	PRO; EST; expressed sequence tag; PCR primer; hybridisation;
probe;	blood coagulation disorder; cancer; cellular adhesion disorder;
secreted	protein; transmembrane protein; ss.
Homo	sapiens.
WO9946281-A2.	

PD



Db 317 acaaggggaaaggttgccttggaatcaccctggagaaagcgaagcaggtactgtct 376  
 QY 361 gtgatttggaagatcccgaaatttggtagaactggtatagtagtattgcccgcctca 420  
 Db 377 gtgatttggaagatcccgaaatttggtagaactggtatagtagtattgcccgcctca 436  
 QY 421 agacatctatgaatttgcagaagtgtgtagcagggatagggaactgaagagaat 480  
 Db 437 agacatctatgaatttgcagaagtgtgtagcagggatagggaactgaagagaat 496  
 QY 481 tctactacatcgtcaggaagagaactacaggaatccctaacccactgcaggaatc 540  
 Db 497 tctactacatcgtcaggaagagaactacaggaatccctaacccactgcaggaatc 556  
 QY 541 ggggtggaatcgtacatgcctcccaaggatgaagtcgcaacactcatcgtgactatg 600  
 Db 557 ggggtggaatcgtacatgcctcccaaggatgaagtcgcaacactcatcgtgactatg 616  
 QY 601 ttgcaagagtggtcttcttcgggtgttcattgacgtgaatgacactgaaaggaggac 660  
 Db 617 ttgcaagagtggtcttcttcgggtgttcattgacgtgaatgacactgaaaggaggac 676  
 QY 661 agtactatgtccagacacactccactcagacactatagcaactgaatgagggggaac 720  
 Db 677 agtactatgtccagacacactccactcagacactatagcaactgaatgagggggaac 736  
 QY 721 ccagcgacctatgttcagtgaggtgtgtgagatgctgagctctggcagatggaatg 780  
 Db 737 ccagcgacctatgttcagtgaggtgtgtgagatgctgagctctggcagatggaatg 796  
 QY 781 acacagagtgccatctaccatgtacitctgtctgagttcatcaagaagaaagtaac 840  
 Db 797 acacagagtgccatctaccatgtacitctgtctgagttcatcaagaagaaagtaac 856  
 QY 841 ttccctcatcagctattgtctatttctctgtgacgtcattacattatgttatcca 900  
 Db 857 ttccctcatcagctattgtctatttctctgtgacgtcattacattatgttatcca 916  
 QY 901 tcttttttctgattgactactacattgtatctgagtcacacatagctagaaaaatgctaaa 960  
 Db 917 tcttttttctgattgactactacattgtatctgagtcacacatagctagaaaaatgctaaa 976  
 QY 961 ctgaggtatgagctcccatcata 985  
 Db 977 ctgaggtatgagctcccatcata 1001

RESULT 3  
 IN AAC78480  
 XX AAC78480 standard; cDNA; 1016 BP.  
 XX AAC78480;  
 DT 88-FEB-2001 (first entry)  
 XX  
 DE Human PRO702 (UNQ366) nucleotide sequence SEQ ID NO:96.  
 XX  
 XX Human: secreted protein; transmembrane protein; PRO; EST; cytosolic;  
 XX expressed sequence tag; detection; cancer; ss.  
 XX Homo sapiens.  
 XX WO200053756 A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 18-FEB-2000; 2000WO-US04341.  
 XX  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 29-MAR-1999; 99US-0126773.  
 PR 21-APR-1999; 99US-0130232.

PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 26-JUL-1999; 99US-0145898.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX (GETH ) GENENTECH INC.  
 XX  
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;  
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;  
 PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;  
 XX  
 DR WPI: 2000-611143/58.  
 DR P-PSDB; AAB44264.  
 XX  
 PT Novel PRO polypeptides and polynucleotides used in detection methods,  
 PT to target bioactive molecules to specific cells, and to modulate  
 PT cellular activities -  
 XX  
 PS Claim 2: Fig 36; 61pp; English.  
 XX  
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
 CC sequence tag) sequences which encode secreted or transmembrane PRO  
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
 CC activity. The PRO polynucleotides and polypeptides can be used for detecting  
 CC the presence of PRO polypeptides in samples, for linking bioactive  
 CC molecules to cells and for modulating biological activities of cells,  
 CC using the polypeptides for specific targeting. The polypeptide targeting  
 CC can be used to kill the target cells, e.g. for the treatment of cancers.  
 CC The polypeptide pairs provide specific targeting of bioactive molecules  
 CC to cells. AAC78600 to AAC7987 represent PCR primers and probes used in  
 CC the isolation of the PRO polynucleotide sequences.  
 XX  
 SQ Sequence 1016 BP; 312 A; 197 C; 261 G; 246 T; 0 other;  
 Query Match 61.5%; Score 980.2; DB 21; Length 1016;  
 Best Local Similarity 99.7%; Pred. No. 4.9e-278;  
 Matches 982; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 cagcaatgaatggcttgcctccttgcctcgaagaaacccaatttatctctgtactat 60  
 Db 17 cagcaatgaatggcttgcctccttgcctcgaagaaacccaatttatctctgtactat 76  
 QY 61 tctttttgcaattcagagctggtgtggtatgtagcgcgtaccgtgaagtct 120  
 Db 77 tctttttgcaattcagagctggtgtggtatgtagcgcgtaccgtgaagtct 136  
 QY 121 gtgccacacacacattccacagaccacccaagagagatggtgtaaaaggagatccag 180  
 Db 137 gtgccacacacacattccacagaccacccaagagagatggtgtaaaaggagatccag 196  
 QY 181 gagaagagggaagcagcgaagtgaggacgcatggcgccgaaggaattaaaggagac 240  
 Db 197 gagaagagggaagcagcgaagtgaggacgcatggcgccgaaggaattaaaggagac 256  
 QY 241 tgggtgatgagcagcggggaataattggaagactggcgccatggggaagaggtg 300  
 Db 257 tgggtgatgagcagcggggaataattggaagactggcgccatggggaagaggtg 316  
 QY 301 acaaggggaaaaaggttcttggaaatcttggaatacttggaagaaaggcaagcaggtactgtct 360

BAKE  
pan J  
N-PSDB; AA  
WP: 2001-  
03-2802

DI 8 10:13:15 2002

us-09-600-932-2.rag

SEQ 1D NO: 2

File C:\tm  
Page 5  
AC NO: AAY79510, Database: A-Genesys-  
03-2802

XX PD 14-SEP-2000.  
XX FF 06-JAN-2000; 2000WO-US00376.  
XX PR 08-MAR-1999; 99WO-US05028.  
XX PR 02-JUN-1999; 99WO-US12252.  
XX PR 23-JUN-1999; 99US-0141037.  
XX PR 07-JUL-1999; 99US-0143048.  
XX PR 26-JUL-1999; 99US-0145698.  
XX PR 30-NOV-1999; 99WO-US28313.  
XX PR 20-DEC-1999; 99WO-US30911.  
XX PR 05-JAN-2000; 2000WO-US00219.  
XX PA (GETH ) GENENTECH INC.  
XX FI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;  
XX PI Watanabe CK, Wood WI;  
XX DR WP1: 2000-572070/53.  
XX DR N-PSDB; AAC58385.  
XX FI Thirty PRO polynucleotides encoding PRO polypeptides, useful in the  
XX PT treatment, diagnosis and prevention of cancer -  
XX XX Claim 61; Fig 38; 286pp; English.  
XX CC The present invention describes an isolated antibody that binds to  
XX CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,  
XX CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,  
XX CC PRO1025, PRO1030, PRO1037, PRO1107, PRO1153, PRO1182, PRO1184,  
XX CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,  
XX CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell  
XX CC growth. The PRO polypeptides and nucleotides are useful in the  
XX CC treatment, diagnosis and prevention of cancer. The antibodies and other  
XX CC anti-tumour compounds may be used to treat various conditions, including  
XX CC those characterised by overexpression and/or activation of the amplified  
XX CC PRO genes. Exemplary conditions or disorders to be treated with such  
XX CC antibodies and other compounds include benign or malignant tumours  
XX CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,  
XX CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic  
XX CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),  
XX CC leukaemias and lymphoid malignancies, other disorders such as neuronal,  
XX CC glial, astrocytic, hypothalamic and other glandular, macrophagal,  
XX CC epithelial, stromal and blastocoele disorders, and inflammatory,  
XX CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR  
XX CC primers and hybridisation probes used in the isolation of the human PRO  
XX CC sequences. AAC58367 to AAC58396 and AAC24057 to AAC24089 represent human  
XX CC PRO polynucleotide and protein sequences given in the exemplification of  
XX CC the present invention.  
XX XX Sequence 271 AA;  
XX SQ

Query Match 48.3%; Score 716.5; DB 21; Length 271;  
Best Local Similarity 47.4%; Pred. No. 3e-64;  
Matches 129; Conservative 60; Mismatches 76; Indels 7; Gaps 3;  
QY 8 LRNFQILLVFLILQISGLDIDSRPTA-EVCATHNISPFGKDGKGDGPGEEKHCK 66  
DB 1 mrgnlavglvlsiaflslpsghpqpagddacsqilvqglkgdagekd--kgatgr 57  
QY 67 VGRMGPKGKIGELGDMGDRNI---GKTGPTRKKGDKGKQLLGPGEKAGTVCDCQR 123  
DB 58 pgrvgtgkdmqdkgsvgrhkgipgskgkgsdgdpdpngpelpcecsq 117  
QY 124 YRKEVQLDISIARLTKSMFKVKNVAGIRETEKFFYIVQEKYRESLTHCRIGML 183  
DB 118 lraigemnqvqsltselkfknagavreteskiyvlvkeekydaqlscqgggl 177  
QY 184 AMPKDEAANTLADYVAKSFFRVFVIGNDLREGQYMTDTNPLQNSNNWNGESDPY 243  
DB 178 smpkdeanglmaaylaqaglarvfigndlekegafvysdhspsmrtnwrsgeppnny 237

QY 244 GHEDCVELSSGRWNTDECHLTMYFVCEFIKK 275  
DB 238 deedcvemvsggwndvachtmyimcefdke 269  
RESULT 6  
ID AAY79510  
XX ID AAY79510 standard; Protein; 271 AA.  
XX AC AAY79510;  
XX DT 01-AUG-2000 (first entry)  
XX DE Human carbohydrate-associated protein CRBAP-6.  
XX KW CRBAP-6; carbohydrate-associated protein 6; human;  
XX KW autoimmune disorder; inflammation; gastrointestinal disorder;  
XX KW infection; reproductive disorder; neurological disorder;  
XX KW eye disorder; cell proliferation; cancer; diagnosis; therapy.  
XX OS Homo sapiens.  
XX FH Key  
FT Peptide 1..20  
FT Protein 21..271  
FT Modified-site 90 /note= "O-phosphorylated"  
FT Modified-site 154 /note= "O-phosphorylated"  
FT Domain 41..112 /note= "collagen-like domain"  
FT Domain 247..256 /note= "C-type lectin domain"  
FT Domain 44..77 /note= "C1q domain protein sequence"  
FT Domain 150..265 /note= "C-type lectin family sequence"  
XX WO2000018922-A2.  
XX 06-APR-2000.  
XX 29-SEP-1999; 99WO-US2685.  
XX 01-OCT-1998; 98US-0164785.  
XX 06-OCT-1998; 98US-0167179.  
XX 13-NOV-1998; 98US-0191838.  
XX 03-DEC-1998; 98US-0205656.  
XX 03-DEC-1998; 98US-2223333.  
XX (INCY-) INCYTE PHARM INC.  
XX AU-Young J, Lal P, Bandman O, Reddy R, Baughn MR, Yue H;  
XX Hillman JL;  
XX WPI: 2000-317516/27.  
XX N-PSDB; AA234946.  
XX Novel carbohydrate-associated proteins used for the prevention and  
XX treatment of autoimmune/inflammatory disorders of e.g. the  
XX gastrointestinal and reproductive systems -  
XX Claim 1; Page 88; 104pp; English.  
XX The present sequence is that of a novel human carbohydrate-associated  
XX protein, termed CRBAP-6, as deduced from cDNA (see AA294946) obtained  
XX from adrenal tumour cDNA library ADRETT06. CRBAP-6 has chemical  
XX and structural similarity to bovine lung surfactant protein D (32%  
XX identity). CRBAP-6 is expressed in liver, kidney, ovary, gut,  
XX adrenal gland and secretory epithelium tissues. The invention



Date is not good

AC NO: AAY41698, Database: A-Genes, 89-032802

CC human collectin protein and its encoding polynucleotide. The human  
CC collectin exhibits antibacterial and antiviral activity and can be used  
CC as an agent for the treatment of human bacterial and viral infections.  
XX This sequence represents the novel human collectin.  
SQ Sequence 277 AA;

Query Match 100.0%; Score 1484; DB 20; Length 277;  
Best Local Similarity 100.0%; Pred No. 4.7e-142;  
Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNGFASLRNQFILLVLELLQISGLDSDSRTAEVCATHTISPPKGDGDEKDPGE 60  
DQ 1 MNGFASLRNQFILLVLELLQISGLDSDSRTAEVCATHTISPPKGDGDEKDPGE 60  
QY 61 EGKHKVGKMGPKCKGELGMDGRNGIKGTGPIGKKGDKGKGLLIPGKKGAGTVD 120  
DQ 61 EGKHKVGKMGPKCKGELGMDGRNGIKGTGPIGKKGDKGKGLLIPGKKGAGTVD 120  
QY 121 CGRYKFKVGLDISIARLAKSMKFKVKNVIAGIRETEKFFYIVQEKNYRESLTHCRIRG 180  
DQ 121 CGRYKFKVGLDISIARLAKSMKFKVKNVIAGIRETEKFFYIVQEKNYRESLTHCRIRG 180  
QY 181 GMLAMPKDEAANTLIADYVAKSGFRVFIQVNDLREGQYFTDNTPLQYNSWNEGEPS 240  
DQ 181 GMLAMPKDEAANTLIADYVAKSGFRVFIQVNDLREGQYFTDNTPLQYNSWNEGEPS 240  
QY 241 DPYGHEDCEVMSLGGWNTCHLTMVFCFFKXKK 277  
DQ 241 DPYGHEDCEVMSLGGWNTCHLTMVFCFFKXKK 277

RESULT 2

RAY41698  
ID RAY41698 standard; Protein; 277 AA.

AC RAY41698;

DT 07-DEC-1999 (first entry)

DE Human PR0702 protein sequence.

KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
KW secreted protein; transmembrane protein.

OS Homo sapiens.

PN WO9946281-A2.

PD 16-SEP-1999.

PF 08-MAR-1999; 99WO-US05028.

PR 10-MAR-1998; 98US-0077450.

PR 11-MAR-1998; 98US-0077632.

PR 11-MAR-1998; 98US-0077641.

PR 11-MAR-1998; 98US-0077649.

PR 12-MAR-1998; 98US-0077791.

PR 13-MAR-1998; 98US-0078004.

PR 17-MAR-1998; 98US-0040220.

PR 20-MAR-1998; 98US-0078886.

PR 20-MAR-1998; 98US-0078910.

PR 20-MAR-1998; 98US-0078936.

PR 20-MAR-1998; 98US-0078939.

PR 25-MAR-1998; 98US-0079294.

PR 26-MAR-1998; 98US-0079656.

PR 27-MAR-1998; 98US-0079663.

PR 27-MAR-1998; 98US-0079669.

PR 27-MAR-1998; 98US-0079728.

PR 27-MAR-1998; 98US-0079786.

PR 30-MAR-1998; 98US-0079920.  
PR 30-MAR-1998; 98US-0079923.  
PR 31-MAR-1998; 98US-0080105.  
PR 31-MAR-1998; 98US-0080107.  
PR 31-MAR-1998; 98US-0080165.  
PR 31-MAR-1998; 98US-0080194.  
PR 01-APR-1998; 98US-0080327.  
PR 01-APR-1998; 98US-0080328.  
PR 01-APR-1998; 98US-0080333.  
PR 01-APR-1998; 98US-0080334.  
PR 08-APR-1998; 98US-0081049.  
PR 08-APR-1998; 98US-0081070.  
PR 08-APR-1998; 98US-0081071.  
PR 09-APR-1998; 98US-0081195.  
PR 09-APR-1998; 98US-0081203.  
PR 09-APR-1998; 98US-0081229.  
PR 15-APR-1998; 98US-0081817.  
PR 15-APR-1998; 98US-0081838.  
PR 15-APR-1998; 98US-0081952.  
PR 15-APR-1998; 98US-0081955.  
PR 21-APR-1998; 98US-0082568.  
PR 21-APR-1998; 98US-0082569.  
PR 22-APR-1998; 98US-0082700.  
PR 22-APR-1998; 98US-0082704.  
PR 22-APR-1998; 98US-0082804.  
PR 23-APR-1998; 98US-0082766.  
PR 27-APR-1998; 98US-0083336.  
PR 28-APR-1998; 98US-0083322.  
PR 29-APR-1998; 98US-0083392.  
PR 29-APR-1998; 98US-0083495.  
PR 29-APR-1998; 98US-0083496.  
PR 29-APR-1998; 98US-0083499.  
PR 29-APR-1998; 98US-0083500.  
PR 29-APR-1998; 98US-0083545.  
PR 29-APR-1998; 98US-0083554.  
PR 29-APR-1998; 98US-0083558.  
PR 30-APR-1998; 98US-0083559.  
PR 05-MAY-1998; 98US-0083742.  
PR 06-MAY-1998; 98US-0084414.  
PR 06-MAY-1998; 98US-0084441.  
PR 07-MAY-1998; 98US-0084598.  
PR 07-MAY-1998; 98US-0084600.  
PR 07-MAY-1998; 98US-0084627.  
PR 07-MAY-1998; 98US-0084637.  
PR 07-MAY-1998; 98US-0084639.  
PR 07-MAY-1998; 98US-0084640.  
PR 07-MAY-1998; 98US-0084643.  
PR 13-MAY-1998; 98US-0085323.  
PR 13-MAY-1998; 98US-0085338.  
PR 13-MAY-1998; 98US-0085339.  
PR 15-MAY-1998; 98US-0085573.  
PR 15-MAY-1998; 98US-0085579.  
PR 15-MAY-1998; 98US-0085580.  
PR 15-MAY-1998; 98US-0085582.  
PR 15-MAY-1998; 98US-0085689.  
PR 15-MAY-1998; 98US-0085697.  
PR 15-MAY-1998; 98US-0085700.  
PR 15-MAY-1998; 98US-0085704.  
PR 18-MAY-1998; 98US-0086023.  
PR 22-MAY-1998; 98US-0086392.  
PR 22-MAY-1998; 98US-0086414.  
PR 22-MAY-1998; 98US-0086430.  
PR 22-MAY-1998; 98US-0086486.  
PR 28-MAY-1998; 98US-0087098.  
PR 28-MAY-1998; 98US-0087106.  
PR 28-MAY-1998; 98US-0087208.  
PR 30-JUL-1998; 98US-0094651.  
PR 11-SEP-1998; 98US-0100038.

(GETH) GENENTECH INC.

XX

XX

PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;  
 XX WPI; 1999-551358/46.  
 DR N-PSDB; AA233973.  
 XX  
 PT New secreted and transmembrane polypeptides and their polynucleotides,  
 PT useful for treating blood coagulation disorders, cancers and cellular  
 PT adhesion disorders.  
 XX  
 PS Claim 12; Fig 37; 530pp; English.  
 XX  
 CC The present invention describes secreted and transmembrane polypeptides  
 CC and their polynucleotides. The nucleotide sequences are useful as  
 CC sources of probes, primers, for chromosome mapping, and for generation  
 CC of antisense sequences. They can also be used to create transgenic  
 CC animals. The proteins can be used to treat a variety of diseases and  
 CC disorders, depending on their function. Diseases that may be treated  
 CC include blood coagulation disorders, cancers and cellular adhesion  
 CC disorders. They may also be used to raise antibodies. AA233891 to  
 CC AA234338, and AA41685 to AA41774 represent polynucleotide and  
 CC polypeptide sequence given in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 277 AA;

Query Match 99.2%; Score 1472; DB 20; Length 277;  
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;  
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MNGFASLLRNQFILLVFLQLQISGLDIDSRTAEVCATHTISPGKGDGKGDPE 60  
 Db 1 mgfaslrrnqfllvflqlqisglidsrtaevcathtispqkdgdekdpge 60  
 QY 61 EGKHGKVGKMGPKIGKELGMDGRNIGTKTGPIGKKGKGLLIGIPGKAGTVC 120  
 Db 61 egkhgkvgrmpkgikgelgmdgdnigtktpigkkgkgligipgkagtvcd 120  
 QY 121 CGRYKFKVQGLDISIARLTKSMKFKVKNVIAGIRETEEFYIVQEKNYRESLTHCRIRG 180  
 Db 121 cgryrkfvqglidisarltsmkfkvnviagireteekfyivqeknyreslthcrirg 180  
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRFVIGVNDLREGQYMTNTPLQYNSWNEGEPS 240  
 Db 181 gmlampkdeaanliadyvaksgffrfvignvdleregymtdntplqynswnegeps 240  
 QY 241 DPYGHEDCEVMSLSSGRWNDECHLTMYFVCEFIKKKK 277  
 Db 241 dpyghedcvmllssgrwndtechltmyfvcfeikkkk 277

RESULT 3  
 AAB44554  
 ID AAB44254 standard; Protein: 277 AA.  
 AC AAB44254;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human PRO702 (UNQ366) protein sequence SEQ ID NO:97.  
 XX  
 KW Human; secreted protein; transmembrane protein; PRO; EST; cytotstatic;  
 KW expressed sequence tag; detection; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200053756-A2.  
 XX  
 PD 14-SEP-2000;  
 XX  
 PF 18-FEB-2000; 2000WO-US04341.  
 XX  
 PR 08-MAR-1999; 99WO-US05028.

AR 12-MAR-1999; 99US-0123957.  
 PR 29-MAR-1999; 99US-0126773.  
 PR 21-APR-1999; 99US-0130232.  
 PR 28-APR-1999; 99US-0131445.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 05-JAN-2000; 99WO-US31274.  
 PR 06-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 XX  
 PA (GETH) GENENTECH INC.  
 XX  
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;  
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;  
 PI Goddard A, Godowski PJ, Grimaldi C, Gurney AL, Hillan KJ;  
 PI Kljavin LJ, Luo SS, Napier MA, Pan J, Paoni NF, Roy MA;  
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;  
 XX  
 DR WPI; 2000-611443/58.  
 DR N-PSDB; AAC78480.  
 XX  
 PT Novel PRO polypeptides and polynucleotides used in detection methods,  
 PT to target bioactive molecules to specific cells, and to modulate  
 PT cellular activities.  
 XX  
 PS Claim 12; Fig 37; 636pp; English.  
 XX  
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
 CC sequence tag) sequences which encode secreted or transmembrane PRO  
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
 CC activity. The polynucleotides and polypeptides can be used for detecting  
 CC the presence of PRO polypeptides in samples, for linking bioactive  
 CC molecules to cells and for modulating biological activities of cells,  
 CC using the polypeptides for specific targeting. The polypeptide targeting  
 CC can be used to kill the target cells, e.g. for the treatment of cancers.  
 CC The polypeptide pairs provided specific targeting of bioactive molecules  
 CC to cells. AAC78600 to AAC78898 represent PCR primers and probes used in  
 CC the isolation of the PRO polynucleotide sequences.  
 XX  
 SQ Sequence 277 AA;

Query Match 99.2%; Score 1472; DB 21; Length 277;  
 Best Local Similarity 99.3%; Pred. No. 7.8e-141;  
 Matches 275; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MNGFASLLRNQFILLVFLQLQISGLDIDSRTAEVCATHTISPGKGDGKGDPE 60  
 Db 1 mgfaslrrnqfllvflqlqisglidsrtaevcathtispqkdgdekdpge 60  
 QY 61 EGKHGKVGKMGPKIGKELGMDGRNIGTKTGPIGKKGKGLLIGIPGKAGTVC 120  
 Db 61 egkhgkvgrmpkgikgelgmdgdnigtktpigkkgkgligipgkagtvcd 120  
 QY 121 CGRYKFKVQGLDISIARLTKSMKFKVKNVIAGIRETEEFYIVQEKNYRESLTHCRIRG 180  
 Db 121 cgryrkfvqglidisarltsmkfkvnviagireteekfyivqeknyreslthcrirg 180  
 QY 181 GMLAMPKDEAANTLIADYVAKSGFFRFVIGVNDLREGQYMTNTPLQYNSWNEGEPS 240  
 Db 181 gmlampkdeaanliadyvaksgffrfvignvdleregymtdntplqynswnegeps 240  
 QY 241 DPYGHEDCEVMSLSSGRWNDECHLTMYFVCEFIKKKK 277  
 Db 241 dpyghedcvmllssgrwndtechltmyfvcfeikkkk 277

DR P-PSDB: RAY25518.

XX PT New collectin protein of human origin and DNA encoding it

XX PS Claim 2; Page 39-42; 58pp; Japanese.

XX CC This invention describes the isolation and characterization of a novel human collectin protein and its encoding polynucleotide. The human collectin exhibits antibacterial and antiviral activity and can be used as an agent for the treatment of human bacterial and viral infections.

XX CC This sequence encodes the novel human collectin.

XX SQ Sequence 1595 bp; 444 A; 322 C; 382 G; 447 T; 0 other;

Query Match 100.0%; Score 1595; DB 20; Length 1595;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1595; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagcaatgaatggttgcacccctgttcgaagaacaaatttaccctcctgactat 60  
DB 1 cagcaatgaatggttgcacccctgttcgaagaacaaatttaccctcctgactat 60  
QY 61 tcttttgcaattcagtgctggtggtggtggtggtggtggtggtggtggtggt 120  
DB 61 tcttttgcaattcagtgctggtggtggtggtggtggtggtggtggtggtggt 120  
QY 121 gtgcac 180  
DB 121 gtgcac 180  
QY 181 gagaaggggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 240  
DB 181 gagaaggggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 240  
QY 241 tgggtgatatggagatggagatggagatggagatggagatggagatggagatgg 300  
DB 241 tgggtgatatggagatggagatggagatggagatggagatggagatggagatgg 300  
QY 301 acaaggggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 360  
DB 301 acaaggggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 360  
QY 361 gtgattgtggaagatgacgaagatgacgaagatgacgaagatgacgaagatgacga 420  
DB 361 gtgattgtggaagatgacgaagatgacgaagatgacgaagatgacgaagatgacga 420  
QY 421 agacatctatgaattgtgcaagatgcaagatgcaagatgcaagatgcaagatgca 480  
DB 421 agacatctatgaattgtgcaagatgcaagatgcaagatgcaagatgcaagatgca 480  
QY 481 tctactacatctgaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 540  
DB 481 tctactacatctgaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 540  
QY 541 ggggtggaatgtagcagatgtagcagatgtagcagatgtagcagatgtagcagatg 600  
DB 541 ggggtggaatgtagcagatgtagcagatgtagcagatgtagcagatgtagcagatg 600  
QY 601 ttgccaagagtggttcttcggttcttcggttcttcggttcttcggttcttcggtt 660  
DB 601 ttgccaagagtggttcttcggttcttcggttcttcggttcttcggttcttcggtt 660  
QY 661 agtacatgttcacagac 720  
DB 661 agtacatgttcacagac 720  
QY 721 ccagcgaacccctatggtcagagactgtgtgagatgtgtgagatgtgtgagatgtgt 780  
DB 721 ccagcgaacccctatggtcagagactgtgtgagatgtgtgagatgtgtgagatgtgt 780  
QY 781 acacagatgcatcttaccatgtactttgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840

DB XX

SEQ ID NO:1  
AC NO: AA333973  
Database: N-geneseq-032802

DB 781 acacagatgcatcttaccatgtactttgtgtgtgtgtgtgtgtgtgtgtgtgtgt 840  
QY 841 ttcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 900  
DB 841 ttcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 900  
QY 901 ttccttttttctcctgatttactacatttgcctcagtcacacatagctagaataatgctaa 960  
DB 901 ttccttttttctcctgatttactacatttgcctcagtcacacatagctagaataatgctaa 960  
QY 961 ctgaggtatgagcctccatcatcgtcttcttctgtgtgtgtgtgtgtgtgtgtgtgtgt 1020  
DB 961 ctgaggtatgagcctccatcatcgtcttcttctgtgtgtgtgtgtgtgtgtgtgtgtgt 1020  
QY 1021 ggtatgtattgacccaataactcgcaggtgtacatggtctgtgagagagataatttatt 1080  
DB 1021 ggtatgtattgacccaataactcgcaggtgtacatggtctgtgagagagataatttatt 1080  
QY 1081 actaatttgcaagagatggt 1140  
DB 1081 actaatttgcaagagatggt 1140  
QY 1141 tgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1200  
DB 1141 tgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1200  
QY 1201 actggtgattgattgagaggt 1260  
DB 1201 actggtgattgattgagaggt 1260  
QY 1261 acatgtacaaagggcttctgtgagcaatgataagatcttgaatccaagatgccagatg 1320  
DB 1261 acatgtacaaagggcttctgtgagcaatgataagatcttgaatccaagatgccagatg 1320  
QY 1321 ttttaccagtcacacccctatggtccatggtctacttctgtgaggttctctgtgtgtgtgtgt 1380  
DB 1321 ttttaccagtcacacccctatggtccatggtctacttctgtgaggttctctgtgtgtgtgt 1380  
QY 1381 catgaatgcttttaaccccaagcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440  
DB 1381 catgaatgcttttaaccccaagcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440  
QY 1441 cagaccatgtggaatgataataactcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500  
DB 1441 cagaccatgtggaatgataataactcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500  
QY 1501 tataccaagtagtctgt 1560  
DB 1501 tataccaagtagtctgt 1560  
QY 1561 atatagtcacactttgatttaagaaaaacggagcc 1595  
DB 1561 atatagtcacactttgatttaagaaaaacggagcc 1595

RESULT 2  
AA333973  
ID AA333973 standard; CDNA: 1016 BP.  
XX AC AA333973;  
XX AC AA333973;  
XX AC AA333973;  
DT 07-DEC-1999 (first entry)  
XX Human PRO702 nucleotide sequence.  
XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
XX probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
XX secreted protein; transmembrane protein; ss.  
XX Homo sapiens.  
XX PN W09946281-A2.  
XX



XX	PF	08-MAR-1998;	99WO-US0503028.
XX	XX		
XX	XX		
PR	PR	10-MAR-1998;	98US-0077450.
PR	PR	11-MAR-1998;	98US-0077632.
PR	PR	11-MAR-1998;	98US-0077641.
PR	PR	12-MAR-1998;	98US-0077649.
PR	PR	12-MAR-1998;	98US-0077791.
PR	PR	17-MAR-1998;	98US-0078004.
PR	PR	20-MAR-1998;	98US-0040220.
PR	PR	20-MAR-1998;	98US-0078886.
PR	PR	20-MAR-1998;	98US-0078910.
PR	PR	20-MAR-1998;	98US-0078936.
PR	PR	25-MAR-1998;	98US-0078939.
PR	PR	26-MAR-1998;	98US-0079294.
PR	PR	27-MAR-1998;	98US-0079656.
PR	PR	27-MAR-1998;	98US-0079663.
PR	PR	27-MAR-1998;	98US-0079664.
PR	PR	27-MAR-1998;	98US-0079689.
PR	PR	27-MAR-1998;	98US-0079728.
PR	PR	30-MAR-1998;	98US-0079786.
PR	PR	30-MAR-1998;	98US-0079920.
PR	PR	31-MAR-1998;	98US-0079923.
PR	PR	31-MAR-1998;	98US-0080105.
PR	PR	31-MAR-1998;	98US-0080107.
PR	PR	31-MAR-1998;	98US-0080165.
PR	PR	01-APR-1998;	98US-0080194.
PR	PR	01-APR-1998;	98US-0080327.
PR	PR	01-APR-1998;	98US-0080328.
PR	PR	01-APR-1998;	98US-0080333.
PR	PR	08-APR-1998;	98US-0080334.
PR	PR	08-APR-1998;	98US-0081049.
PR	PR	08-APR-1998;	98US-0081070.
PR	PR	09-APR-1998;	98US-0081071.
PR	PR	09-APR-1998;	98US-0081195.
PR	PR	09-APR-1998;	98US-0081203.
PR	PR	15-APR-1998;	98US-0081229.
PR	PR	15-APR-1998;	98US-0081838.
PR	PR	15-APR-1998;	98US-0081817.
PR	PR	15-APR-1998;	98US-0081952.
PR	PR	21-APR-1998;	98US-0081955.
PR	PR	21-APR-1998;	98US-0082568.
PR	PR	22-APR-1998;	98US-0082569.
PR	PR	22-APR-1998;	98US-0082700.
PR	PR	22-APR-1998;	98US-0082704.
PR	PR	23-APR-1998;	98US-0082804.
PR	PR	23-APR-1998;	98US-0082767.
PR	PR	27-APR-1998;	98US-0082766.
PR	PR	28-APR-1998;	98US-0083336.
PR	PR	29-APR-1998;	98US-0083342.
PR	PR	29-APR-1998;	98US-0083352.
PR	PR	29-APR-1998;	98US-0083495.
PR	PR	29-APR-1998;	98US-0083436.
PR	PR	29-APR-1998;	98US-0083500.
PR	PR	29-APR-1998;	98US-0083545.
PR	PR	29-APR-1998;	98US-0083554.
PR	PR	29-APR-1998;	98US-0083558.
PR	PR	30-APR-1998;	98US-0083742.
PR	PR	05-MAY-1998;	98US-0083742.
PR	PR	06-MAY-1998;	98US-0084366.
PR	PR	06-MAY-1998;	98US-0084414.
PR	PR	07-MAY-1998;	98US-0084441.
PR	PR	07-MAY-1998;	98US-0084598.
PR	PR	07-MAY-1998;	98US-0084600.
PR	PR	07-MAY-1998;	98US-0084627.
PR	PR	07-MAY-1998;	98US-0084637.
PR	PR	07-MAY-1998;	98US-0084639.
PR	PR	07-MAY-1998;	98US-0084640.
PR	PR	13-MAY-1998;	98US-0084643.
PR	PR	13-MAY-1998;	98US-0085328.
PR	PR	13-MAY-1998;	98US-0085333.

PR	13-MAY-1998;	98US-0085339.
PR	15-MAY-1998;	98US-0085573.
PR	15-MAY-1998;	98US-0085579.
PR	15-MAY-1998;	98US-0085580.
PR	15-MAY-1998;	98US-0085582.
PR	15-MAY-1998;	98US-0085689.
PR	15-MAY-1998;	98US-0085697.
PR	15-MAY-1998;	98US-0085700.
PR	15-MAY-1998;	98US-0085704.
PR	18-MAY-1998;	98US-0086023.
PR	22-MAY-1998;	98US-0086392.
PR	22-MAY-1998;	98US-0086414.
PR	22-MAY-1998;	98US-0086430.
PR	22-MAY-1998;	98US-0086486.
PR	28-MAY-1998;	98US-0087098.
PR	28-MAY-1998;	98US-0087106.
PR	28-MAY-1998;	98US-0087208.
PR	30-JUL-1998;	98US-0094651.
PR	11-SEP-1998;	98US-0100038.

(GETH ) GENENTECH INC.

Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J:

WPI; 1999-551358/46.  
P-PSDB; AAY41698.

PT New secreted and transmembrane polypeptides and their polynucleotides,  
PT useful for treating blood coagulation disorders, cancers and cellular  
PT adhesion disorders -

Claim 2; Fig 36; 530pp; English.

CC The present invention describes secreted and transmembrane polypeptides  
CC and their polynucleotides. The nucleotide sequences are useful as  
CC sources of probes, primers, for chromosome mapping, and for generation  
CC of antisense sequences. They can also be used to create transgenic  
CC animals. The proteins can be used to treat a variety of diseases and  
CC disorders, depending on their function. Diseases that may be treated  
CC include blood coagulation disorders, cancers and cellular adhesion  
CC disorders. They may also be used to raise antibodies. AA233991 to  
CC AA234338, and AA41685 to AA41774 represent polynucleotide and  
CC polypeptide sequence given in the exemplification of the present  
CC invention.

Sequence 1016 BP; 312 A; 197 C; 261 G; 246 T; 0 other:

Query Match 61.5%; Score 980.2; DB 20; Length 1016;  
Best Local Similarity 99.7%;  
Matches 982; Conservative 0; Mismatches 2.

	Conserved	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0;
QY	1	cagcaatgaatggcttgcatactctgcttcgagaagaaacaaattatctctcgtgactat	60						
Db	17	cagcaatgaatggcttgcatactctgcttcgagaagaaacaaattatctctcgtgactat	76						
QY	61	ttcttttgcaaattcagatctgggtctggatattgatccctcctaccgtgaagctt	120						
Db	77	ttcttttgcaaattcagatctgggtctggatattgatccctcctaccgtgaagctt	136						
QY	121	gtgcacacacacaatttcacagagcccaagagagatgatgtgaaagaggagatccag	180						
Db	137	gtgcacacacacaatttcacagagcccaagagagatgatgtgaaagaggagatccag	196						
QY	181	gagaagagggaagcatggcaaatggagcgcattgggcccgaagaagaaattaaaggagaaac	240						
Db	197	gagaagagggaagcatggcaaatggagcgcattgggcccgaagaagaaattaaaggagaaac	256						
QY	241	tgggtgatattggagatcggggcaaatattggcaagactggcccatgggaagaagggtg	300						
Db	257	tgggtgatattggagatcaggggcaaatattggcaagactgggcccattgggaagaagggtg	316						
QY	301	acaaaggggaaaaaggtttgcttggaaatccttgagaaaaaggccaaagcagggtactgtct	360						



QY	244	GHENCVEMLSGGWNDTECHLTMYFVCEFIKK	275
Db	238	deedcvmasgsgwvndvachtmyfmcfdke	269
		:             :      :      :             :      :	
RESULT	6		
ID	AAV79510		
XX	AAV79510 standard; Protein; 271 AA.		
AC	AAV79510;		
XX	01-AUG-2000 (first entry)		
DE	Human carbohydrate-associated protein CRBAP-6.		
XX	CRBAP-6; carbohydrate-associated protein 6; human;		
KW	autoimmune disorder; inflammation; gastrointestinal disorder;		
KW	infection; reproductive disorder; neurological disorder;		
KW	eye disorder; cell proliferation; cancer; diagnosis; therapy.		
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..20	
FT	Protein	/note= "signal peptide"	
FT	Modified-site	21..271	
FT	Modified-site	/note= "mature protein"	
FT	Modified-site	90	
FT	Modified-site	/note= "O-phosphorylated"	
FT	Domain	154	
FT	Domain	/note= "O-phosphorylated"	
FT	Domain	41..112	
FT	Domain	/note= "collagen-like domain"	
FT	Domain	247..256	
FT	Domain	/note= "C-type lectin domain"	
FT	Domain	44..77	
FT	Domain	/note= "C1q domain protein sequence"	
FT	Domain	150..265	
FT	Domain	/note= "C-type lectin family sequence"	
PN	WC2000018922-A2.		
PD	06-APR-2000.		
XX	29-SEP-1999;	99WO-US22685.	
XX	01-OCT-1998;	98US-014785.	
XX	06-OCT-1998;	98US-0167179.	
XX	13-NOV-1998;	98US-0191838.	
XX	03-DEC-1998;	98US-0205656.	
XX	03-DEC-1998;	98US-2223333.	
PA	(INCY-) INCYTE PHARM INC.		
XX	Au-Young J, Lal P, Bandman O, Reddy R, Baughn MR, Yue H;		
PI	Hillman JL;		
PI	WPI: 2000-317516/27.		
DR	N-PSDB; AA294946.		
XX	Novel carbohydrate-associated proteins used for the prevention and		
PT	treatment of autoimmune/inflammatory disorders of e.g. the		
PT	gastrointestinal and reproductive systems -		
XX	Claim 1; Page 88; 104pp; English.		
XX	The present sequence is that of a novel human carbohydrate-associated		
CC	protein, termed CRBAP-6, as deduced from cDNA (see AA294946) obtained		
CC	from adrenal tumour cDNA library ADRETUT06. CRBAP-6 has chemical		
CC	and structural similarity to bovine lung surfactant protein D (32%		
CC	identity). CRBAP-6 is expressed in liver, kidney, ovary, gut,		
CC	adrenal gland and secretory epithelium tissues. The invention		

